

UNIVERSIDADE DE LISBOA  
FACULDADE DE CIÊNCIAS  
DEPARTAMENTO DE MATEMÁTICA

ISCTE- BUSINESS SCHOOL  
INSTITUTO UNIVERSITÁRIO DE LISBOA  
DEPARTAMENTO DE FINANÇAS



**Ciências**  
**ULisboa**

**ISCTE**  **IUL**  
Instituto Universitário de Lisboa

**AVALIAÇÃO EMPÍRICA DO RISCO DE MERCADO:  
ESTIMAÇÃO DO *VALUE-AT-RISK* PELA TEORIA DOS  
VALORES EXTREMOS**

**TATIANA CRISTINA SOARES GARÇÃO**

**Mestrado em Matemática Financeira**

Dissertação orientada por:  
Prof.<sup>a</sup> Doutora Diana Aldea Mendes

2017



## AGRADECIMENTOS

Em primeiro lugar, quero agradecer de um modo muito particular e especial à Prof<sup>a</sup> Dr.<sup>a</sup> Diana Mendes pelo conhecimento científico que partilhou comigo e por toda a dedicação, disponibilidade, compreensão e orientação que teve durante todas as fases deste trabalho, mesmo durante os momentos mais atribulados mostrou-se incansável para que este objetivo fosse concluído.

Agradeço, também, ao Prof. Dr. João Pedro Nunes que sempre me incentivou a prosseguir este trabalho.

Por fim, agradeço aos meus pais, ao meu irmão e aos meus colegas de mestrado por todo o apoio, ajuda e incentivo que me foram transmitindo ao longo deste trabalho.

## RESUMO

Nos últimos anos, os mercados financeiros têm apresentado comportamentos que se têm traduzido em perdas avultadas em especial para as instituições financeiras. Nesse sentido, os órgãos reguladores têm fomentado a implementação de metodologias de prevenção e gestão de risco.

Ao nível das métricas mais populares para medir o risco encontram-se atualmente o *Value-at-Risk* (VaR), contudo as metodologias tradicionais de cálculo do VaR supõem normalidade e acomodam mal as ocorrências extremas da distribuição dos retornos. Pois é conhecido o fato de que distribuições de retornos das séries financeiras costumam apresentar caudas mais pesadas do que as de uma distribuição normal.

O objetivo principal deste trabalho é realçar a importância da Teoria de Valores Extremos (EVT) no cálculo do *Value-at-Risk* (VaR), investigar a possibilidade das hipóteses exigidas pelas metodologias tradicionais, e a influência que as características tipicamente encontradas em séries financeiras exercem sobre os modelos de cálculo do VaR mais utilizados. Estuda-se, também, a possibilidade de se obter uma melhoria significativa do ponto de vista do controlo do risco, através da utilização da EVT.

Foram aplicadas técnicas de *backtesting*, como os testes de Kupiec (1995) e de Christoffersen (1998), na avaliação da performance dos diferentes modelos de previsão do VaR

São apresentados de uma forma sucinta alguns dos principais resultados ligados à teoria dos valores extremos e também são apresentadas algumas estatísticas que possibilitam a simplificação do processo de reconhecimento de dados de cauda pesada. A modelação da cauda é um assunto de particular interesse, dá-se mais importância a dois métodos de modelação da cauda.

**PALAVRAS-CHAVE:** *Value-at-Risk*, GARCH, Riskmetrics, *Extreme Value Theory*, *Generalized Pareto Distribution*, *Block Maxima*, *peak over threshold*

## ABSTRACT

In the last years, financial markets have shown some behaviors that resulted on huge losses, especially for financial institutions. In that context, regulators have encouraged the implementation of preventive and management risk methodologies.

On the top of most popular metrics to measure risk, we found nowadays Value-at-Risk (VaR), however the traditional VaR calculation methodologies assume normality and hardly accommodate the extreme occurrences of the distribution of returns. Actually, it is quite well known that distributions of returns from the financial series tend to have heavier tails than those of a normal distribution.

The main purpose of this dissertation is to enhance the importance of Extreme Value Theory (EVT) in the calculation of Value-at-Risk (VaR), investigate the possibility of the hypotheses required by traditional methodologies, and the influence that the characteristics typically found in financial series on the most widely used VaR calculation models. The possibility of achieving a significant improvement from the point of view of risk control through the use of EVT is also studied.

Backtesting technics like Kupiec tests (1995) and Christoffersen (1998) have been applied in the performance evaluation of the different models for forecasting VaR.

A short introduction to the main results inherent in the EVT and also, a set of statistics to simplify the recognition process of heavy tailed data is provided. Tail modelling is a subject of particular interest in this dissertation, hence two methods of tail modeling are receiving more importance.

**KEYWORDS:** *Value-at-Risk, GARCH, Riskmetrics, Extreme Value Theory, Generalized Pareto Distribution, Block Maxima, peak over threshold*

## ÍNDICE:

INTRODUÇÃO .....	1
<b>CAPÍTULO I</b>	
VALUE-AT-RISK (VAR) .....	3
1.1 Risco Financeiro .....	3
1.2 Value at Risk (VaR) .....	4
1.3 Expected Shortfall (ES) .....	4
1.4 Estimação do VaR .....	5
1.4.1 Identificação dos parâmetros .....	5
1.4.2 VaR para distribuições gerais .....	5
1.4.3 VaR para distribuições paramétricas .....	6
1.4.4 Modelo Basileia II .....	7
1.4.5 Condições normais e valores extremos .....	9
1.4.6 Black Swans .....	9
1.5 PSI-20 .....	10
1.6 Revisão bibliográfica .....	11
<b>CAPÍTULO II</b>	
MODELOS ECONOMETRÍCOS PARA A ESTIMAÇÃO DO RISCO .....	13
2.1 Séries Temporais .....	13
2.1.1 Modelo ARMA .....	17
2.1.2 Modelo GARCH .....	18
2.2 Métodos Paramétricos e Não-Paramétricos .....	20
2.3 Backtesting .....	21
2.3.1 Teste de taxa de falha .....	22
2.3.2 Teste de Kupiec e Christofferson .....	22
2.4 Extreme Value Theory (EVT) .....	25
2.4.1 Distribuição dos valores máximos .....	26
2.4.2 The Peak Over Threshold Method .....	29
<b>CAPÍTULO III</b>	
MODELOS ESTIMADOS/ESTIMAÇÃO E VALIDAÇÃO DO VAR .....	35
3.1 Análise descritiva e testes de raiz unitária .....	35
3.2 Estimação do VaR pelos modelos paramétricos e não-paramétricos .....	37
3.3 Estimação do VaR pelos modelos EVT .....	39
CONCLUSÃO .....	44
BIBLIOGRAFIA .....	47
ANEXOS .....	49
Anexo 1: Teste de raiz unitária à série PSI-20 para o teste ADF .....	49
Anexo 2: Teste de raiz unitária à série PSI-20 para o teste PP .....	50
Anexo 3: Teste de raiz unitária à série dos retornos para o teste ADF .....	51
Anexo 4: Teste de raiz unitária à série dos retornos para o teste PP .....	52

## ÍNDICE DE TABELAS

TABELA 1.1. VALORES CRÍTICOS $Z\alpha$ .....	7
TABELA 1.2 .....	8
TABELA 1.3. COMPOSIÇÃO DO ÍNDICE EM 13 DE JUNHO DE 2017: .....	10
TABELA 3.1. TESTES DE RAIZ UNITÁRIA .....	36
TABELA 3.2. RESULTADOS DOS MODELOS NA PREVISÃO DO VAR PARA UMA PREVISÃO DE 5 DIAS .....	38
TABELA 3.3. RESULTADOS DOS MODELOS NA PREVISÃO DO VAR PARA UMA PREVISÃO DE 1 DIA .....	39
TABELA 3.4. PERCENTAGEM DE FALHA .....	39

## ÍNDICE DE FIGURAS

FIGURA 1.1. EXPECTED SHORTFALL (ES) E VAR .....	4
FIGURA 1.2. VALORES CRÍTICOS $Z_{\alpha}$ .....	7
FIGURA 2.1 (B). SÉRIE NÃO ESTACIONÁRIA.....	15
FIGURA 2.1 (A). SÉRIE ESTACIONÁRIA .....	15
FIGURA 2.2. AUTOCORRELAÇÃO DOS RESÍDUOS PADRONIZADOS .....	20
FIGURA 2.4. DISTRIBUIÇÃO DA FUNÇÃO $F$ E DISTRIBUIÇÃO CONDICIONAL DA FUNÇÃO $F_u$ (FONTE MANFRED GILLI E EVIS KÄELLEZI (2006)) .....	29
FIGURA 2.5. DISTRIBUIÇÃO GENERALIZADA DE PARETO $G_{\xi}, \sigma x$ PARA $\sigma=1$ (FONTE MANFRED GILLI E EVIS KÄELLEZI (2006)) .....	30
FIGURA 3.1. ANÁLISE GRÁFICA E HISTOGRAMA DA SÉRIE PSI-20.....	36
FIGURA 3.2. ANÁLISE GRÁFICA E HISTOGRAMA DA SÉRIE DOS RETORNOS DO PSI-20 .....	36
FIGURA 3.3. OUTPUT DA ESTIMAÇÃO /PREVISÃO DO VAR POR VÁRIOS MÉTODOS.....	37
FIGURA 3.4. ANÁLISE GRÁFICA PARA OS MODELOS GARCH.....	38
FIGURA 3.5. ANÁLISE GRÁFICA DA GPD .....	40
FIGURA 3.6. ANÁLISE GRÁFICA DA DISTRIBUIÇÃO DOS VALORES MÁXIMOS .....	41
FIGURA 3.7. ANÁLISE GRÁFICA DOS RESÍDUOS .....	41
FIGURA 3.8. MEAN EXCESSO PLOT.....	42
FIGURA 3.9. GRÁFICO DA FUNÇÃO HILL .....	42
FIGURA 3.10. MODELO GPD AJUSTADO COM AS EXCEDÊNCIAS ACIMA DO VALOR DE THRESHOLD .....	43
FIGURA 3.11. Q-Q PLOT DOS RESÍDUOS.....	43
FIGURA 3.12. ESTIMAÇÃO DO QUANTIL .....	44



## INTRODUÇÃO

No mercado financeiro, conseguir saber qual o valor do risco a que uma instituição está sujeita é um desafio constante, uma vez que a sua má avaliação e a consequente não mitigação pode levar a que a instituição tenha grandes prejuízos. Assim, surge a necessidade de identificar o que é o risco financeiro, como medi-lo, como modelá-lo e, principalmente, como prevê-lo. Pois o conhecimento antecipado deste valor é cada vez mais fundamental para que, de forma rápida e exata, existam diretrizes para as tomadas de decisão necessárias.

O conceito de “risco financeiro” pode ser apresentado como sendo a incerteza de retorno de um investimento perante um possível acontecimento futuro e incerto, independente do investidor e que poderá originar prejuízos. Holton (2004) defendeu que o risco envolve duas componentes essenciais: exposição e incerteza.

Tem surgido uma vasta discussão relativamente à forma de medir o risco.

Em 1994, o banco de investimento J. P. Morgan tornou pública a metodologia *Riskmetrics*, dando uma contribuição importante para a difusão dos modelos de gestão do risco. Com esta nova metodologia, o cálculo do VaR foi adotado como sendo o melhor método para quantificar o risco de mercado e tem sido usado até aos dias de hoje. No entanto, este não apresenta sensibilidade aos valores extremos e tem sido criticado por não ser uma medida de risco coerente.

Assim, o principal propósito deste trabalho será verificar que metodologias tradicionais do cálculo do VaR têm melhor desempenho no cálculo do risco associado ao principal índice de ações português, o PSI-20, e comparar com o estudo da modelação do VaR baseada na Teoria de Valores Extremos (EVT).

A base de dados é constituída por 2892 observações diárias, 5 dias por semana, obtidas entre Janeiro de 2006 e Janeiro de 2017. De uma forma mais pormenorizada, o VaR irá ser calculado através de modelos GARCH(1,1).

Cada um destes modelos será depois sujeito a uma análise dos resultados através dos testes propostos por Kupiec (1995) e Christoffersen (1998) e que permitirá perceber qual o modelo mais adequado.

Para o estudo da Teoria dos Valores Extremos irão ser considerados dois métodos mais conhecidos, o método dos blocos de máximos e o método *peak over threshold* (POT), para tal irá ser utilizado o software MATLAB.



# Capítulo I

## *Value-at-Risk* (VaR)

### 1.1 Risco Financeiro

O risco é a probabilidade de ocorrência de eventos que afetem, neste caso, de forma significativa a condição financeira de uma instituição, independentemente da existência ou da qualidade dos controles instituídos. Resulta exclusivamente de fatores exógenos, ou seja, fatores externos que estão fora do poder de intervenção de uma instituição.

A noção de risco financeiro refere-se às hipóteses do resultado de uma operação vinculada às finanças não seja o previsto. Estes riscos estão relacionados com eventuais perdas associadas às atividades financeiras. Quanto maior o risco financeiro, maiores as possibilidades de o resultado ser diferente do esperado.

Os riscos podem ser de diferentes tipos:

- ✓ **Risco de Crédito:** é a possibilidade de que uma das partes de um contrato relacionado com uma operação de empréstimo não cumpra com as suas obrigações.
- ✓ **Risco de Mercado:** é a probabilidade de ocorrência de impactos negativos nos resultados ou no capital, devido a movimentos desfavoráveis no preço de mercado dos instrumentos da carteira de negociação (flutuações em cotações de ações, preços de mercadorias, taxas de juro, taxas de câmbio, ...).
- ✓ **Risco de Taxa de Juro:** está associado com a volatilidade que lhes é inerente. Esta volatilidade tem a ver com as variações que as taxas de juro sofrem, motivadas por fatores económicos, sociais, políticos ou outros, como as condições do mercado, a política das instituições financeira ou a qualidade creditícia do agente solicitador de crédito. As taxas de juro podem ser fixas ou variáveis.
- ✓ **Risco Cambial:** está associado à variação do preço de uma divisa relativamente a outra e “ (...) pode ser gerado por atividades de comércio (exportação/importação) com o estrangeiro, por uma atividade financeira em divisas e também pelo desenvolvimento multinacional da empresa.” Gillot e Pion (1993: p.17).
- ✓ **Risco de Liquidez:** é assumido pelas empresas financeiras que outorgam créditos e que precisam de dispor de dinheiro em efetivo (líquido) de maneira permanente para poder operar.

Ao longo desta tese vai-se estimar o risco financeiro, para tal duas das formas mais utilizadas são o VaR e o ES, estudados mais à frente.

## 1.2 Value at Risk (VaR)

O VaR tem como medida objetiva procurar reportar a probabilidade de ocorrer uma perda máxima esperada associada a um horizonte temporal, recorrendo a valores de mercado.

O VaR é a perda máxima esperada que, em condições normais de mercado e num dado horizonte temporal (1 dia, 10 dias, ...), se pode verificar com um determinado nível de significância  $\alpha$  (=5%, 1%, ...) ou de confiança  $(1-\alpha)$  (=95%, 99%, ...).

Em termos estatísticos, o VaR não é mais do que o  $\alpha$  percentil (que se designa por valor crítico) da distribuição de probabilidades das variações do valor da carteira no horizonte temporal determinado. Assim, o VaR para um nível de confiança de  $(1 - \alpha)$  é dado pela seguinte expressão (o  $\alpha$ -Quantile Risk Metric):

$$P[\Delta V < -VaR] = \alpha \quad (1.1)$$

sendo,  $V$  uma variável aleatória que representa a variação ( $\Delta V$ ) no valor da carteira no horizonte temporal determinado.

## 1.3 Expected Shortfall (ES)

Uma medida complementar e estreitamente relacionada ao VaR é o valor médio da perda quando excede o quantil  $\alpha$ , isto é, o expected shortfall:

$$ES_\alpha = \frac{1}{1-\alpha} \int_0^{1-\alpha} VaR_\gamma(P) d\gamma, \quad (1.2)$$

onde  $VaR_\gamma$  é o valor do risco para o nível de confiança  $\gamma$ , que muda de 0 para  $\alpha$ .

O *Expected Shortfall* (ES), em contraste com o VaR, fornece informações sobre as perdas, que ocorrem quando o nível de confiança é excedido e, portanto, pode potencialmente avaliar as perdas extremas na cauda de distribuição.

Na figura 1.1 está ilustrado um gráfico com as duas medidas de risco referidas anteriormente, tendo sido considerado um  $\alpha = 5\%$ .

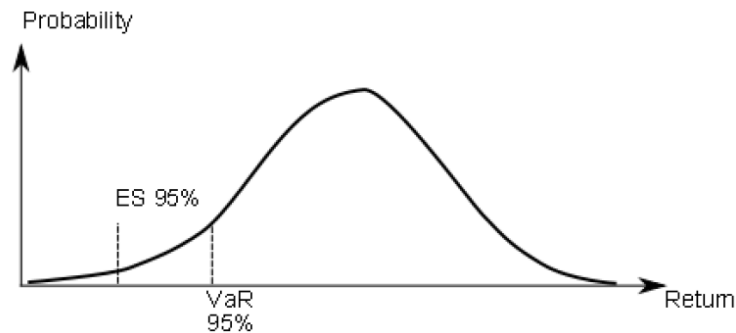


Figura 1.1. *Expected Shortfall* (ES) e VaR

## 1.4 Estimação do VaR

Na estimação do VaR os valores são enviesados por condições normais e por valores extremos (*Black Swan*), estudados mais à frente.

Este cálculo exige identificação/estimação de um conjunto de parâmetros/informação:

- ✓ Valor da carteira de ativos e identificação dos fatores de risco;
- ✓ Nível de significância ( $\alpha$ ) ou de confiança ( $1 - \alpha$ );
- ✓ Horizonte temporal ( $h$ );
- ✓ Distribuição de probabilidades.

Matematicamente, o VaR pode ser resumido como,

$$VaR_{\alpha} = F_L(l) \geq \alpha \quad (1.3)$$

que simplesmente coloca estados de que o VaR é o  $\alpha$  quantil de alguma distribuição de perdas  $F_L$ .

Embora o VaR seja uma maneira intuitiva de resumir o risco, ainda é um modelo e, como tal, faz, em certa medida, simplificação excessiva da estimativa de risco.

### 1.4.1 Identificação dos parâmetros

#### Nível de confiança

Deverá refletir o nível de aversão ao risco e a magnitude dos custos a incorrer numa perda superior ao VaR. Geralmente utiliza-se 95%, 97.5% ou 99%.

O *JPMorgan Riskmetrics*<sup>1</sup> considera um nível de confiança de 95%, enquanto que o Comité de Basileia define o nível de confiança em 99%.

#### Horizonte Temporal

É definido em função da liquidez da carteira e da respetiva estratégia de gestão. Não deverá ser inferior ao tempo necessário para liquidar o portfolio em condições normais de mercado.

Geralmente utiliza-se um horizonte temporal diário (bancos), semanal ou mensal (fundos de pensões, fundos de investimento).

O *JPM Riskmetrics* assume um horizonte temporal diário, enquanto que o Comité de Basileia define o horizonte temporal em 10 *trading days*.

#### Distribuição de probabilidades

Existem duas abordagens:

Distribuição empírica: consiste em não assumir uma distribuição teórica de probabilidades para a taxa de rendibilidade

Distribuição paramétrica: assume que a taxa de rendibilidade segue uma distribuição normal

### 1.4.2 VaR para distribuições gerais

Para efeitos de cálculo do VaR, assume-se, apenas, que as variações do valor da carteira são independentes e identicamente distribuídas (i.i.d.).

A determinação do VaR pode ser efetuada a partir de um histograma das variações do valor da carteira (distribuição empírica), num dado horizonte temporal, correspondendo o valor em risco ao  $\alpha$  percentil amostral.

<sup>1</sup> Devido à crescente atenção ao risco, em outubro de 1994, J.P. Morgan lançou *Riskmetrics*™. Uma metodologia de estimativa do risco de mercado que se baseia na experiência de gestão do risco de mercado da Morgan, acompanhada por conjuntos de dados de volatilidade e correlação cobrindo os principais mercados

### 1.4.3 VaR para distribuições paramétricas

Enquanto, que a abordagem empírica apenas utiliza o *ranking* das observações, a abordagem paramétrica utiliza informação sobre toda a distribuição.

Esta abordagem baseia-se normalmente na distribuição normal, pois:

- ✓ As taxas de rendibilidade dos ativos apresentam distribuição aproximadamente normal;
- ✓ A distribuição da média amostral converge para a normal com o aumento do número de observações;
- ✓ A distribuição de probabilidade da rendibilidade é caracterizada por apenas dois parâmetros (média e variância).

#### Distribuição Normal

Utilizam-se taxas de rendibilidade em regime de capitalização contínua (rendibilidade geométrica) em vez de taxas de rendibilidade em regime de capitalização discreta (rendibilidade aritmética) por apresentarem duas vantagens:

- ✓ Se a taxa de rendibilidade geométrica tem distribuição normal, o preço do ativo e o valor da carteira seguem uma distribuição lognormal;
- ✓ A taxa de rendibilidade, num dado intervalo de tempo, é igual à soma das taxas de rendibilidade em subintervalos desse mesmo período, pelo que a taxa de rendibilidade no intervalo de tempo tem também distribuição normal.

Geralmente utiliza-se a distribuição da taxa de rendibilidade geométrica (capitalização contínua) da carteira em vez da distribuição da variação do valor da carteira,

$$Prob [R < R^*] = \alpha$$

Sendo o  $R^*$  igual a:  $R^* = \mu + Z_\alpha \sigma$

onde,

$Z_\alpha$  = valor da distribuição normal cumulativa standardizada ( $z$ );

$\alpha$  = desvio padrão da taxa de rendibilidade da carteira;

$\mu$  = taxa de rendibilidade esperada da carteira.

#### VaR (zero)

Explicitar o VaR face ao valor atual de mercado do portfólio: muito usual em horizontes temporais diários,

$$VaR (zero) = V_0 R^* = V_0 (\mu + z_\alpha \sigma)$$

#### VaR (mean)

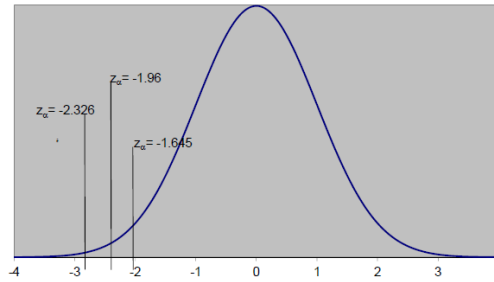
Reportar o VaR em relação a um valor esperado do portfólio com base numa taxa de rendibilidade esperada,

$$VaR (mean) = V_0 [R^* - \mu] = V_0 z_\alpha \sigma$$

Na tabela 1.1 estão indicados os valores críticos  $Z_\alpha$  (número de desvios padrão à esquerda da média) mais frequentemente utilizados para reporte do VaR:

Tabela 1.1. Valores críticos  $Z_\alpha$ 

$1-\alpha$	99%	97.5%	95%	90%
$Z_\alpha$	-2.326	-1.960	-1.645	-1.282

Figura 1.2. Valores críticos  $Z_\alpha$ 

Para calcular o VaR, é necessário estimar os parâmetros  $\mu$  e  $\sigma^2$  da taxa de rendibilidade com base numa amostra aleatória das respetivas observações históricas ( $R_1, R_2, \dots, R_N$ ).

Sabe-se que o estimador centrado para a média da população é dado pela média amostral:

$$\bar{R} = \sum_{i=1}^N \frac{R_i}{N} \quad (1.4)$$

Assim como se sabe que o estimador centrado para a variância da população é dado pela variância amostral:

$$s^2 = \sum_{i=1}^N \frac{(R_i - \bar{R})^2}{N - 1} \quad (1.5)$$

#### 1.4.4 Modelo Basileia II

O modelo Basileia II é um modelo que cumpre as recomendações mínimas do Basileia II. O modelo que será referido como o modelo de Basileia II, é um modelo bastante simples, que será usado como um modelo de referência. Embora simples, esta configuração e os pressupostos subjacentes são amplamente utilizados, tanto na prática como na teoria descrevendo o modelo do VaR. O Basileia recomenda usar um horizonte temporal de 10 dias ao estimar o risco de mercado que indiretamente leva em consideração o período de liquidação necessário para os gestores de portfólio, quando ocorrem mudanças no mercado.

Embora o modelo de Basileia II não dê uma estimativa precisa do risco futuro medido pela taxa de falha observada, a precisão pode ser melhorada significativamente pela distinção entre mercados normais e extremos.

##### 1.4.4.1 Pressupostos

O Modelo Basileia II recomenda calcular VaR num intervalo de confiança de 99% e com um horizonte temporal de um dia aumentado para 10 dias, usando a raiz quadrada do tempo. Os dados necessários para calcular o VaR são pelo menos um ano de observações. Espera-se que um ano seja igual a 252 dias de negociação.

O Basileia não fornece diretrizes diretas sobre como estimar o risco. Uma abordagem comum é usar o  $\sigma$  como uma estimativa para a volatilidade.

Embora o Basileia não ofereça diretrizes específicas quanto à escolha da distribuição, é comum na prática e na teoria assumir que os retornos seguem uma distribuição gaussiana. Ao modelar o risco de crédito, este pode ser uma boa aproximação devido ao horizonte temporal relativamente longo.

No entanto, na gestão do risco de mercado, o risco é modelado num horizonte mais curto, com o objetivo de estimar o risco de curto prazo. Assumindo que os retornos seguem o teorema do limite central, não serão necessariamente uma distribuição de Gauss num curto horizonte, embora, ao longo

de um horizonte temporal longo, os retornos possam ser aproximadamente iguais e independentemente distribuídos.

Sendo que  $\alpha$  descreve o nível de significância de uma distribuição normal de Gauss e  $\pi(i)$  o valor da carteira no dia  $i$ , onde  $i = 1$  é a primeira observação após um período inicial de 252 dias usado para a estimativa inicial de  $\hat{\sigma}(i)$ .

Sendo que  $\sigma(i-1)_T$  descreve a volatilidade diária calculada ao longo dos últimos  $T$  dias no instante  $i-1$  e assumindo que  $\hat{\sigma}(i) = \sigma(i-1)_T$ . Por outras palavras, a volatilidade nos últimos  $T$  dias é uma estimativa da volatilidade diária futura.

$\tau$  é definido como o fator de tempo das escalas VaR para o horizonte temporal desejado. Seguindo o teorema do limite central, os retornos são assumidos de forma independentes e identicamente distribuídos (i.i.d.). Na tabela 1.2, são apresentados os pressupostos por trás do modelo de Basileia II, onde o  $r$  são as séries de retorno.

Tabela 1.2 . Pressupostos do modelo Basileia II

Estimativa para o risco futuro	$\hat{\sigma}(i) = \sigma(i-1)_T$
Distribuição dos retornos	$r \sim N(\mu, \sigma)$
Pressuposto implícito	$r - i. i. d$

O VaR é definido como,

$$VaR(i)_\alpha = \pi(i) \cdot z_\alpha \cdot \hat{\sigma}(i) \cdot \tau \quad (1.6)$$

Seguindo as recomendações de Basileia II utiliza-se um intervalo de confiança de 99%, para a distribuição Gauss e um ano de observações para estimar  $\alpha$ , assim, o modelo pode ser reescrito como,

$$\begin{aligned} VaR(i)_{0.99} &= \pi(i) \cdot z_\alpha \cdot \hat{\sigma}(i) \cdot \tau \\ &= \pi(i) \cdot 2.33 \cdot \hat{\sigma}(i) \cdot \sqrt{10} \end{aligned} \quad (1.7)$$

A equação dá-nos o VaR diário de 10 dias para um nível de confiança de 99% multiplicando o valor do portfólio pelo desvio padrão diário futuro estimado com base nos retornos observados no ano passado, dimensionados pela raiz quadrada do tempo. Isto será referido como o modelo de Basileia II.

Uma violação do VaR é quando,

$$\begin{aligned} VaR(i)_\alpha &< \frac{P}{L(i)} \\ &\text{onde,} \\ \frac{P}{L(i)} &= \pi(i-\tau) - \pi(i) \end{aligned} \quad (1.8)$$

O uso da notação de 10 dias para o VaR diário ilustra que o *profit/loss* no instante  $i(P/L(i))$  e é calculado usando uma *rolling window*, o que tecnicamente significa que a perda (ganho) nos últimos dias é calculada usando observações sobrepostas, indiretamente recomendadas em Basileia II.



### 1.4.5 Condições normais e valores extremos

"Os modelos VaR funcionam em mercados normais, e os mercados normais são definidos como aqueles em que os modelos funcionam", Christopher C. Finger, April, 2009. Esta definição de mercados normais é claramente inútil. Pois, é necessária uma outra precisão mais prática, antes que os mercados possam ser divididos em padrões e extremos.

Assumindo que os retornos seguem uma distribuição de Gauss, os mercados agora experimentam grandes retornos medidos em valores absolutos. Por exemplo, 5% dos retornos serão superiores a aproximadamente  $\pm 2$  desvios padrão da média. Por isso, a probabilidade de observar um desses retornos dentro de um determinado período de duas semanas (10 dias) será

$$\binom{n}{x} p^x (1-p)^{n-x} \quad (1.9)$$

$$=$$

$$\binom{10}{1} 0.05^1 \cdot 0.095^9 = 31.51\%$$

assumindo que os retornos são i.i.d. Da mesma forma, a probabilidade de ver dois retornos tão grandes em duas semanas seria então de 7.46%, o que ainda é razoavelmente plausível. A probabilidade de ver três ou mais desses retornos é inferior a 1.05% e, no que se segue, duas semanas em que ocorrem três ou mais dessas observações, serão categorizadas como um período de mercado extremo.

A função abaixo indica como encontrar esses retornos, que são mais de dois desvios padrão da média. A média é aqui estimada como  $\hat{\mu}(i) = \mu(i-1)_{252}$ .

$$I_i = 1_{\mu(i)_{252} - F(0.975)^{-1} \cdot \sigma(i-1)_{252} < r(i) < \mu(i)_{252} + F(0.975)^{-1} \cdot \sigma(i-1)_{252}}$$

Quando menos de três dessas ocorrências são observadas dentro de um período de 10 dias, (isto é, quando  $\sum_{i=1}^{i+9} I_i < 3$ , onde  $i = 1; 10; 19; \dots; n-9$ ), as observações são categorizadas como normais e, assim são incluídas na subamostra normal.

Quando  $\sum_{i=1}^{i+9} I_i \geq 3$ , os dados desse período específico são removidos do conjunto de dados normais e, em vez disso, incluídos na subamostra extrema.

Com base nestas duas subamostras, o VaR normal e extremo pode então ser calculado usando a equação,

$$VaR(i)_{0.99} = \pi(i) \cdot z_{\alpha} \cdot \hat{\sigma}(i) \cdot \tau \quad (1.10)$$

### 1.4.6 Black Swans

Os *Black Swans* são considerados como uma ocorrência inesperada, ao mesmo tempo que têm um grande impacto, neste caso, nos mercados financeiros.

Do mesmo modo, os participantes do mercado racionalizam os *Black Swans* ao fornecer explicações para essas ocorrências depois de terem ocorrido.

Taleb (2007) argumenta que a arrogância epistêmica, que é a incapacidade de conhecer nossas próprias limitações, nos leva a pensar que grandes inesperados choques podem ser previstos na próxima vez. No entanto, como os *Black Swans* ocorrem em relação às nossas próprias expectativas, e como tendemos a racionalizá-las, isso faz-nos pensar que o mundo é menos aleatório do que realmente é.

Se essa questão não for abordada, isso pode ter sérias consequências, pois os gestores de risco tenderão a acreditar cegamente nas suas próprias habilidades e modelos (limitados), fazendo com que eles permaneçam sem justificção quando ocorrer outro *Black Swan*.

Posto isto, pode-se perguntar como é que isto nos ajudará a prever o risco de amanhã. A verdade é que não pode, isto apenas ajuda-nos a perceber que os modelos têm limitações e que esse conhecimento nos torna muito melhores do que quando acreditávamos cegamente neles.

O principal interesse de Taleb (2007) é fazer com que todos consigamos ver que o mundo é muito mais aleatório do que aquilo em que acreditamos. Podemos reconhecer esses *Black Swans*, mas, ao mesmo tempo, podemos argumentar que existem formas e métodos que podem, até certo ponto, ajudar a indicar em que tipo de mercado estamos e onde estamos indo.

## 1.5 PSI-20

É o principal índice da Euronext Lisboa e o principal índice de referência do mercado de capitais português. É composto pelas ações das vinte maiores empresas cotadas na bolsa de valores de Lisboa e reflete a evolução dos preços dessas ações, que são as de maior liquidez entre as negociadas no mercado português.

O PSI-20 foi lançado com as duas finalidades:

- ✓ Servir de indicador da evolução do mercado acionista português;
- ✓ Servir de suporte à negociação de contratos de futuros e opções

A capitalização bolsista das emissões que compõem o PSI-20 é ajustada pelo *free float*, não podendo cada emissão ter uma ponderação superior a 20% nas datas de revisão periódica da carteira. Devido às suas características, o índice PSI-20 foi selecionado pelo mercado para servir de subjacente a produtos estruturados, cuja rentabilidade depende, de uma ou de outra forma, do comportamento do mercado bolsista português.

Tabela 1.3. Composição do índice em 13 de junho de 2017:

Nome	Sector	Cotação	Volume
Altri	Produtos industriais gerais	4,388	499.032
BCP	Bancos	0,2391	69.480.678
Corticeira Amorim	Bens de consumo	12,43	48.013
CTT	Transporte Industrial	5,634	609.759
EDP	Eletricidade	3,212	3.597.968
EDP Renováveis	Eletricidade	6,973	52.655
Galp Energia	Produção de Petróleo e Gás	13,5	380.261
Ibersol	Viagens e lazer	14,4	5.497
Jerónimo Martins	Retalho alimentar e de medicamentos	17,835	240.424
Mota-Engil	Construção e materiais	2,402	1.049.181
NOS	Media	5,41	411.040
Novabase	Software e serviços informáticos	3,224	30.975
PHAROL	Holding	0,299	4.212.867
REN	Eletricidade	2,858	494.179
Semapa	Sevicultura e papel	16,925	49.885
Sonae Capital	Serviços Financeiros	0,847	330.373
Sonae SGPS	Retalho alimentar e de medicamentos	0,96	3.855.438
The Navigator Company	Sevicultura e papel	4	1.321.344
Unidades Participação Montepio	Bancos	0,454	21.369

## 1.6 Revisão bibliográfica

Este subcapítulo tem como objetivo fornecer uma visão geral da literatura atual sobre o *Value-at-Risk* (VaR) no campo da gestão de riscos e a *Extreme Value Theory* (EVT). O foco é definido no desenvolvimento da metodologia e nas principais fontes de informação sobre VaR como ferramenta de gestão de risco para profissionais de risco.

O documento técnico sobre VaR, lançado pela *Riskmetrics* (1994), tornou-o disponível para o público e rapidamente esta medida ganhou popularidade entre os profissionais de gestão de risco. A adoção generalizada ainda não ocorreu até que o trabalho fundamental de P. Jorion em 1997 surgiu. Até hoje, foram publicadas três edições do livro, cada uma incorporando os últimos desenvolvimentos no campo de gestão de riscos. Desde a sua primeira publicação, ela permaneceu como uma das fontes mais abrangentes de metodologia VaR e estabeleceu-se como um manual de analista do risco. Os pontos principais indicados são que o VaR é uma ferramenta excelente, que melhora a capacidade dos gestores para avaliar os riscos nas diferentes categorias, ao mesmo tempo que fornece um valor simples que pode capturar os riscos de assumir posições e decisões financeiras complexas. O autor analisa as dificuldades potenciais no cálculo do VaR tanto para fins de gestão de risco quanto de regulação.

Houve numerosos estudos e discussões sobre se o VaR é capaz de entregar o que promete. Pritsker (1997) examina a precisão das estimativas do VaR para derivados usando métodos de simulação de Monte Carlo e encontra pouca precisão para opções "*out-of-the-money*". O estudo enfatiza a importância do *trade-off* entre precisão e tempo de computação, bem como a importância de especificar corretamente a distribuição dos choques de fatores.

Daniélsson (2002) examina as limitações dos modelos de risco modernos tanto para a gestão de riscos quanto para regulação. Em particular, ele estuda a precisão dos modelos em diferentes níveis de confiança e descobre que em termos de precisão, os modelos GARCH e *Riskmetrics* são os melhores em 95%, mas o seu desempenho diminui em 99%.

Na prática, os modelos GARCH de baixa ordem, como o GARCH (1,1), são os mais usados, pois são adequados para a maioria das aplicações. McNeil, *et al.*, (2005b), McNeil e Frey (2000), Nyström e Skoglund (2002b), McNeil *et al.* (2005b) e Kuester *et al.* (2006) evidenciam a superioridade de uma abordagem combinada. A distribuição residual dos processos GARCH é, portanto, modelada por uma combinação da distribuição empírica e colunas paramétricas estimadas com base na teoria do valor extremo.

McNeil e Frey (2000) identificam duas características dos métodos baseados na teoria do valor extremo que os tornam particularmente atrativos para a estimativa da cauda (*tail estimation*): primeiro esses métodos são baseados numa teoria estatística sólida e, em segundo lugar, fornecem as caudas de uma distribuição numa forma paramétrica.

No livro sobre previsão de risco financeiro, Daniélsson (2011) combinou a metodologia do VaR com a versatilidade e o poder de processamento do *software* moderno (R e MATLAB). O guia abrangente sobre a implementação de modelos de previsão de risco é complementado com uma análise detalhada da teoria. As questões de precisão e interpretação são abordadas nos artigos anteriores do autor. O autor enfatiza a importância dos pressupostos feitos para cada modelo e a interpretação crítica dos resultados. Para o cálculo do VaR de alta confiança, a metodologia EVT é usada para calcular o

VaR nas regiões extremas das caudas de distribuição. Essa abordagem mitiga a subestimação do risco, uma vez que a distribuição dos retornos nas caudas é melhor capturada pela EVT.

A EVT fornece modelos paramétricos simples para capturar as caudas extremas de uma distribuição e prever riscos. Existem dois métodos amplos de aplicação da EVT: o primeiro é baseado nas distribuições de valor extremo das distribuições de Gumbel, Fréchet ou Weibull que são generalizadas como *Generalized extreme value distribution* (GEV) e conhecidas como *Block Maxima* (Minima) (BMM), o segundo baseia-se na *Generalized Pareto Distribution* (GPD) mais conhecido como a abordagem *peak over threshold* (POT). Os modelos BMM são os mais tradicionais dos dois, e a abordagem BMM adapta-se a um bloco de máximos ou de mínimos (eventos extremos) numa série de dados de observações independentes e identicamente distribuídas (i.i.d). A GEV usa diferentes métodos estatísticos mas o mais comum é o Maximum Likelihood Estimation (MLE). Contudo, o POT é considerado o modelo mais eficiente na modelagem de dados limitados (Gilli e Këllezi, 2006, McNeil, Frey e Embrecht, 2005), uma vez que se encaixa nas superações de um determinado limite num conjunto de dados para GPD e, portanto, não depende tanto do requisito de grande conjuntos de dados como a BMM.

Martins (2015) testou o *Expected Shortfall* (ES) recorrendo à simulação histórica, simulação estocástica (de Monte Carlo), modelos ARCH/GARCH, EWMA e GenericPred (teoria de caos). Para tal, foram utilizadas as séries temporais das cotações diárias do EUR/CHF e EUR/GBP ou os seus retornos diários. Constatou-se que as séries das cotações são irregulares e não estacionárias e que as séries dos retornos, apesar de serem estacionárias, não têm distribuição normal e evidenciam *clusters* de volatilidade (heterocedasticidade).

Segundo Vicente (2012), o EVT oferece conclusões interessantes quando aplicado ao mundo do desporto, mesmo num contexto de análise simplificado. A lição mais importante é a clara nitidez entre abordagens semi-paramétricas e o método POT paramétrico. Em ambos os Estudos de Caso, os resultados e estimativas obtidos são muito semelhantes, para as abordagens acima mencionadas. Pelo contrário, o método *Block Maxima* permanece na margem, produzindo resultados particulares, diferentes dos outros métodos.

## Capítulo II

### Modelos Econométricos para a Estimação do Risco

#### 2.1 Séries Temporais

A teoria das séries temporais tem sido dominada, até recentemente, por modelos lineares, apesar de inúmeras vezes o comportamento das séries económicas sugerir o contrário.

As relações económicas, muitas vezes, podem ser não-lineares:

- ✓ Podem existir efeitos assimétricos sobre as variáveis, assim como nos ciclos económicos, onde os comportamentos são distintos nas contrações e expansões;
- ✓ As variáveis económicas podem ter os seus valores limitados, não podendo aumentar ou diminuir além de um determinado limite, ou seja, constatar-se a existência de *floors* e *ceilings* na variação da série.

Uma série temporal (sucessão cronológica) é um conjunto de observações tomadas em instantes de tempo determinados com intervalos iguais, isto é

$$y_{t_1}, \dots, y_{t_n}, \text{ sendo que } t_1 = t_2 = \dots = t_n \in T$$

e denota-se por  $\{y_t, t \in T\}$ .

O coeficiente de correlação de Pearson de duas variáveis aleatórias  $X$  e  $Y$ , é definido por,

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}. \quad (2.1)$$

Sabe-se que se verifica autocorrelação quando  $X = Y$ .

Quando se representa graficamente o valor obtido por cada um dos coeficientes de autocorrelação (entre -1 e 1), obtém-se o correlograma. A sequência de coeficientes de autocorrelação designa-se por função de autocorrelação (ACF).

A função de autocorrelação (ACF) para a série  $\{y_t, t \in T\}$  define-se por,

$$\rho(t,s) = \frac{Cov(y_t, y_{t-s})}{\sqrt{Var(y_t)Var(y_{t-s})}} = \frac{\gamma(t,s)}{\sqrt{\gamma(0)\gamma(0)}} = \frac{\gamma(t,s)}{\gamma(0)} \quad (2.2)$$

onde  $s$  é o desfasamento da série.

Se a correlação é estável ao longo do tempo, então a ACF toma a forma,

$$\rho(t,s) = \rho(s) = \frac{\gamma(s)}{\gamma(0)}, \forall t \in T \quad (2.3)$$

Verifica-se que,

$$\rho(0) = \frac{\gamma(0)}{\gamma(0)} = 1$$

ou seja, cada série é perfeitamente correlacionada com ela própria.

O coeficiente de autocorrelação parcial de ordem  $s$  denota-se por  $p(s)$  e é definido pelo coeficiente de  $y_{t-s}$  no seguinte modelo linear,

$$y_t = a + a_1 y_{t-1} + \dots + a_{s-1} y_{t-s+1} + r(s) y_{t-s} + \varepsilon_t \quad (2.4)$$

onde  $\varepsilon_t$  não é correlacionado com  $y_{t-1}, \dots, y_{t-s}$ .

O coeficiente mede a intensidade da relação entre duas observações da série, controlando (mantendo constante) o efeito das mesmas,

$$\begin{aligned} y_t &= a_0 + a_1 y_{t-1} + \varepsilon_t \rightarrow p(1) = a_1 \\ y_t &= a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t \rightarrow p(2) = a_2 \end{aligned}$$

Ou seja, a função de autocorrelação parcial (PACF)  $p(s)$  é a correlação entre  $y_t$  e  $y_{t-s}$ , mantendo todos os  $y_h$  constantes para  $t-s < h < t$

Os coeficientes de autocorrelação para diferentes desfasamentos temporais de uma variável, podem ser usados para identificar padrões de comportamento expressos por uma série cronológica.

#### Séries estacionárias/não estacionárias

As propriedades dos estimadores dependem do facto de as séries serem ou não estacionárias. Nas séries estacionárias, as técnicas suavizam as irregularidades de curto prazo.

Uma série  $\{y_t, t \in T\}$  diz-se fracamente estacionária (estacionaridade fraca ou em covariância) se para todo o  $t$  e  $(t-s)$  se verificar as seguintes condições,

- ✓ **Média constante:**  $E(y_t) = E(y_{t-s}) = \mu$
- ✓ **Variância constante:**  $E[(y_t - \mu)^2] = E[(y_{t-s} - \mu)^2] = \sigma_y^2 < \infty$
- ✓ **Covariância constante:**  $E[(y_t - \mu)(y_{t-s} - \mu)] = E[(y_{t-s} - \mu)(y_{t-j-s} - \mu)] = \gamma(s)$

Conclui-se assim que uma série é fracamente estacionária se tem a média e a variância constantes no tempo, e a covariância entre valores desfasados da série depende apenas da desfasagem, ou seja, da distância temporal entre eles. Se uma das três propriedades falha, então a série diz-se não-estacionária.

A estacionaridade implica que os  $y_t$  são identicamente distribuídos e que a natureza de qualquer correlação entre termos adjacentes é a mesma ao longo de todos os períodos.

A estacionaridade implica estacionaridade fraca, mas o contrário não é necessariamente verdadeiro (só é verdade para uma distribuição Gaussiana). Quando conseguimos provar a estacionaridade, referimos, em geral, a estacionaridade fraca.

Todos os processos fracamente estacionários têm as funções de autocorrelação e de autocorrelação parcial a convergir para zero, quando o desfasamento de tempo cresce.

Uma série não-estacionária pode ser estabilizada (estacionarizada) fazendo:

- ✓ **Diferenças sucessivas:**  $\Delta y_t = y_t - y_{t-1}$ ;  $\Delta^2 y_t = y_t - 2y_{t-1} + y_{t-2}, \dots$
- ✓ **Diferenças dos logaritmos dos dados:**  $\Delta \log y_t = \log y_t - \log y_{t-1}$

As figuras que se seguem ilustram algumas séries referidas anteriormente, nomeadamente uma série estacionária (figura 2.1 (a)) e uma série não-estacionária (figura 2.1 (b)).

Na figura 2.1 (a) podemos observar uma variação moderada em torno da média, o que se traduz numa série estacionária, ao contrário do que se observa na figura 2.1 (b).

Na figura 2.1 (b) verifica-se que a série não tem uma tendência constante, apresentando um crescimento nas observações ao longo dos anos, tendo-se registado o maior pico entre 141 e 151. A partir desse momento temporal verifica-se um decréscimo da série.

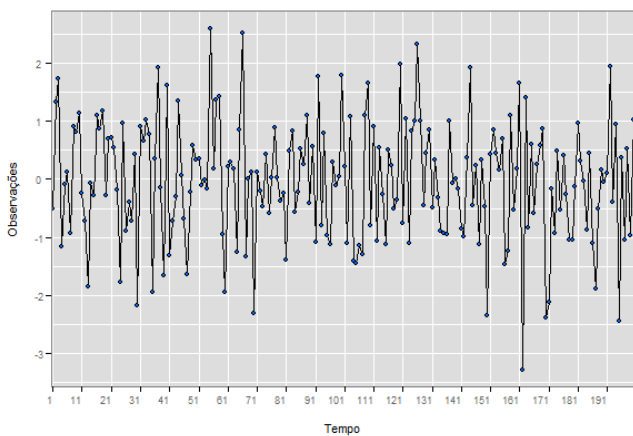


Figura 2.1 (a). Série Estacionária

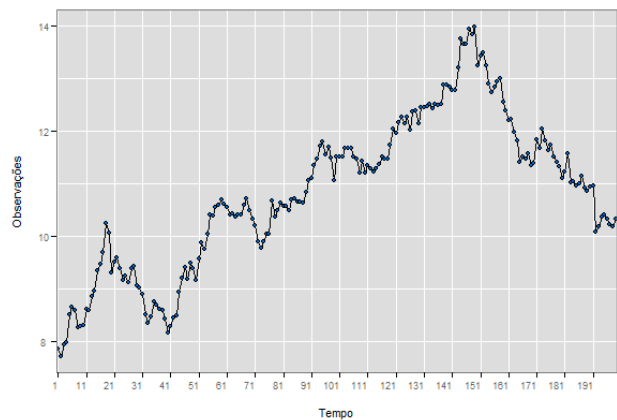


Figura 2.1 (b). Série Não Estacionária

A não-estacionariedade estuda-se em termos estatísticos pelos testes de raiz unitária, pois, estes, podem auxiliar o investigador a proceder a uma correta distinção entre séries cronológicas estacionárias e séries não-estacionárias de forma a evitar o perigo de se obter regressões espúrias na análise. Caso as séries cronológicas sejam não-estacionárias, uma das possíveis transformações com vista a torna-las estacionárias consiste em obter uma série diferenciada da série original, contudo existe perda de informação no processo de diferenciação.

Na maioria dos testes a hipótese nula é de que a série tenha raiz unitária, e portanto não seja estacionária, logo:

$H_0$  : tem raiz unitária (não é estacionária)

$H_1$ : não tem raiz unitária (é estacionária)

Seguidamente são apresentados dois dos testes mais usados de raiz unitária, o teste de *Augmented Dickey-Fuller (ADF)* e o teste de *Phillips – Perron (PP)*.

O teste Dickey-Fuller parte da pressuposição que os termos de erros nas equações do próprio teste são identicamente e independentemente distribuídos (i.i.d), isto é, não apresentam autocorrelação. Em função disso, o teste Dickey-Fuller foi ampliado de forma a incorporar defasamentos em relação a variável que está sendo analisada.

O teste de Dickey-Fuller Aumentado é conhecido como teste ADF (*Augmented Dickey-Fuller*) e requer o estudo sobre a seguinte regressão:

$$\Delta y_t = \alpha + \beta t + \delta y_{t-1} + \sum_{i=1}^m \alpha_i \Delta y_{t-i} + \varepsilon_t \quad (2.5)$$

sendo que, o termo  $\alpha$  representa o coeficiente linear,  $\beta$  o coeficiente de tendência linear,  $\delta$  é o coeficiente de presença de raiz unitária e  $m$  é o número de defasamentos tomados na série.

Neste caso a hipótese nula é dada por

$$H_0: \delta = 0 \\ (\delta = p - 1 = 0)^1$$

Enquanto que,

$$H_1: \delta \neq 0$$

Faz-se uma regressão de  $\Delta y_t$  em  $y_{t-1}, \Delta y_{t-1}, \dots, \Delta y_{t+p-1}$  e calcula-se a estatística  $T$  dada por,

$$T = \frac{\widehat{\delta}}{se(\widehat{\delta})} \quad (2.6)$$

onde  $\widehat{\delta}$  é um estimador para  $\delta$  e  $se(\widehat{\delta})$  é um estimador para desvio padrão do erro de  $\delta$ .

Os valores críticos da estatística  $T$  foram tabelados por Dickey e Fuller através de simulação Monte Carlo e variam nos casos de presença somente de intercepto, presença somente de tendência e presença de ambos.

A principal vantagem do teste ADF em relação ao DF é que ao introduzir um número suficiente de defasamentos, garante-se que os resíduos não apresentem autocorrelação. De um modo geral, para se determinar o número ideal de defasamentos utiliza-se algum critério de informação.

O teste de Phillips - Perron, conhecido como teste PP é uma generalização do teste de Dickey - Fuller para os casos em que os erros  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  são correlacionados e, possivelmente, heterocedásticos. A regressão para este teste é,

$$\Delta y_t = \alpha + \beta t + \delta y_{t-1} + \sum_{i=1}^m \alpha_i \Delta y_{t-i} + \varepsilon_t \quad (2.7)$$

Neste caso, a estatística  $Z$  é calculada por,

$$Z = n\widehat{\delta}_n - \frac{n^2\widehat{\sigma}^2}{2s_n^2}(\widehat{\lambda}_n^2 - \widehat{\gamma}_{0,n}) \quad (2.8)$$

onde,

$$\widehat{\gamma}_{j,n} = \frac{1}{n} \sum_{i=1+j}^n r_i r_{i-j} \\ \widehat{\lambda}_n^2 = \widehat{\gamma}_{0,n} + 2 \sum_{j=1}^q \left(1 - \frac{j}{q+1}\right) \widehat{\gamma}_{j,n}$$

---

1  $y_t = \alpha + \beta t + \gamma y_{t-1} \xrightarrow{\text{acrescentando } (-y_{t-1})} y_t - y_{t-1} = \alpha + \beta t + \gamma y_{t-1} - y_{t-1}$

$\rightarrow \Delta y_t = \alpha + \beta t + y_{t-1}(\gamma - 1),$

onde,  $(\gamma - 1) = \delta$ .



$$s_n^2 = \frac{1}{n-k} \sum_{i=1}^n r_i^2$$

em que o  $r_i$  representa o resíduo em  $y_i$  utilizando estimadores de mínimos quadrados,  $k$  é o número de co-variáveis na regressão e  $q$  é o número de defasamentos utilizados para calcular  $\hat{\lambda}_n^2$ .

Considera-se que  $Z$  trata-se de um ajuste na estatística de Dickley - Fuller. Caso o processo seja não correlacionado temos covariâncias nulas e neste caso,  $\hat{\lambda}_n^2 = \widehat{\gamma}_{0,n}$ . Se o processo não for heterocedástico tem-se que  $se(\delta) = 1/n$  e então  $Z$  é dado por,

$$Z = n\hat{\delta} = \frac{\hat{\delta}}{se(\hat{\delta})} \quad (2.9)$$

isto é,  $Z$  é a estatística de Dickley - Fuller e, portanto, tem a mesma distribuição da estatística do teste ADF, calculada por Dickley - Fuller através de simulação de Monte Carlo.

### 2.1.1 Modelo ARMA

ARMA é um modelo utilizado para ajustar equações com base nos seus valores passados e na média da série. ARMA, vem do inglês *autoregressive moving average models*, que significa Modelo auto-regressivo de médias móveis.

Um modelo ARMA é definido por parâmetros  $p$  e  $q$ , onde o primeiro representa o número de termos auto-regressivos e o segundo a ordem de médias móveis.

Este modelo é a junção dos modelos AR (Modelos Autoregressivos) e MA (Modelos de Médias Móveis) e pode ser representado por,

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (2.10)$$

onde  $\varepsilon_t$  é o termo do erro,  $\phi_1, \dots, \phi_p$  são os coeficientes da parte autorregressiva e  $\theta_1, \dots, \theta_q$  são os coeficientes da parte de média móvel.

Um modelo frequentemente utilizado é o ARMA(1,1), ou seja,

$$X_t = \phi X_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}$$

No caso geral, calcula-se a função de autocovariância (f.a.c.v.) por,

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \dots + \phi_p \gamma_{j-p} + \gamma X_\varepsilon(j) - \theta_1 \gamma X_\varepsilon(j-1) - \dots - \theta_q \gamma X_\varepsilon(j-q)$$

onde,  $\gamma X_\varepsilon(j)$  é a covariância cruzada entre  $X_t$  e  $\varepsilon_t$  definida por,

$$\gamma X_\varepsilon(j) = \begin{cases} = 0, j > 0, \\ \neq 0, j \leq 0. \end{cases}$$

Assim, a f.a.c.v fica,

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \dots + \phi_p \gamma_{j-p}, j > q \quad (2.11)$$

e a f.a.c.v é obtida por,

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \dots + \phi_p \rho_{j-p}, j > q. \quad (2.12)$$

Um processo ARMA (p,q) tem f.a.c. infinita, a qual reduz de acordo com exponenciais e/ou ondas sinusoides amortecidas após o "lag"  $p - q$ . Essa observação é importante na identificação do modelo aos dados observados.

### 2.1.2 Modelo GARCH

O Modelo GARCH (*generalized autoregressive conditional heteroskedasticity model*) é usado para estimar a volatilidade. Em comparação com outros modelos de risco, este método oferece estimativas de volatilidade mais dinâmicas. A modelagem de volatilidade passou por um grande desenvolvimento desde a sua introdução por Engle em 1982 e Bollerslev em 1986. Há vários trabalhos que aplicam os modelos GARCH a séries temporárias financeiras. Muitos dos autores propuseram que um GARCH simples seja sempre suficiente para alcançar bons resultados.

Conforme já foi estudado anteriormente, embora os dados empíricos tenham sido divididos numa amostra com condições normais e condições extremas as flutuações ainda ocorrem na subamostra normal.

No entanto, uma vez que o mercado normal é mais estável, em comparação com as condições extremas de mercado, a metodologia GARCH deve produzir resultados muito satisfatórios.

Embora tanto a variância incondicional como a condicional sejam uma medida de volatilidade, estas são bastante diferentes. Para entender o significado da técnica de modelagem GARCH, é importante compreender a diferença entre estas duas variâncias.

A variância incondicional não tem em consideração as atuais tendências do mercado, mas pode ser considerada como uma medida de variância média, enquanto que a variância condicional tem em conta os movimentos do mercado e considera esse ponto de estimativa. Devido a esta diferença, uma medida de risco condicional será superior quando se trata de séries de retorno heterocedásticas.

Conforme já foi observado no capítulo anterior, o pressuposto subjacente aos retornos de serem i.i.d., foi muito irreal devido ao seu comportamento heterocedástico e autocorrelação dos retornos. Por isso, uma estimativa de volatilidade condicional será melhor como uma previsão de risco.

Nas condições normais de mercado, as estimativas VaR poderiam ser melhoradas se forem estudadas por um modelo com uma distribuição com "caudas mais gordas". A distribuição *t-Student* é uma distribuição para os dados empíricos que utiliza uma abordagem MLE (*maximum likelihood estimation*, em português significa estimativa de máxima verosimilhança). Assim, ao usar uma abordagem MLE para a distribuição *t-Student* e incorporar isso no modelo GARCH, pode ser obtida uma estimativa de risco muito melhor, o que leva a que as estimativas VaR sejam melhores e dinâmicas.

Neste seguimento, presume-se que os retornos possam ser escritos da seguinte forma,

$$X_i = \mu_i + \sigma_i Z_i \quad (2.13)$$

onde  $\sigma_i^2$  é a variância condicional e é definida por,

$$\sigma_i^2 = \alpha_0 + \alpha_1 \varepsilon_{i-1}^2 + \beta_1 \sigma_{i-1}^2 \quad (2.14)$$

Nesta última equação (GARCH (1,1)) o termo  $\varepsilon$  é o erro. Esta é a diferença entre o retorno estimado e o retorno real observado.  $\varepsilon$  também é conhecido como os resíduos nos modelos estimados.

Na prática,  $\varepsilon$  é considerado o ruído branco, o que significa que o seu valor esperado é zero, mas é permitido variar em torno da média com um desvio padrão de  $\sigma$ . Isso faz com que siga um processo normal condicional. É devido a esse termo de erro ( $\varepsilon$ ) que o modelo de estimativa GARCH é considerada condicional. A razão para isso é que, quando os mercados experimentam choques grandes e inesperados, estes são calculados por  $\varepsilon$  e têm em conta nos dias seguintes uma estimativa  $\sigma$ .

A partir da última equação, verifica-se ainda que o impacto que o  $\varepsilon$  tem na estimativa  $\sigma$  é influenciada pelo termo  $\alpha_1$ . Do mesmo modo, as estimativas históricas  $\sigma$  são influenciadas pelo  $\beta_1$ . Também o termo  $\alpha_0$  influencia a estimativa do risco futuro.

É possível analisar o erro no modelo GARCH observando os resíduos padronizados, também conhecidos como inovações,  $Z$ .

Estes podem ser calculados assim,

$$Z_i = \varepsilon_i / \sigma_i$$

$Z_i$  são inovações de ruído branco extraídas de alguma função de distribuição subjacente. Traçar o ACF para os resíduos indicará se há alguma autocorrelação. Como os resíduos são assumidos como ruído branco não devem estar correlacionados.

Se não forem feitas premissas com relação à distribuição subjacente de  $X_i$ , assume-se que a variância condicional é condicional ao conjunto de informações. Isto significa que  $X_i$  pode ser estimado utilizando as observações passadas e sua distribuição até  $i - 1$ .

Os parâmetros do modelo GARCH são condicionados por certas restrições,

$$\begin{aligned}\alpha_1 + \beta_1 &< 1 \\ \alpha_0 &> 0 \\ \alpha_1 &\geq 0 \\ \beta_1 &\geq 0\end{aligned}$$

Estas restrições são necessárias para garantir que a variância incondicional seja finita e positiva e também para garantir que a variância condicional seja sempre positiva. Cada parâmetro tem sua própria interpretação de como ele afeta a volatilidade em um modelo GARCH.

- ✓  $\alpha_1$  é o parâmetro de erro do GARCH e mede a reação da volatilidade condicional aos choques de mercado. Isto significa que ter um grande  $\alpha_1$  implicará uma volatilidade muito sensível aos choques de mercado.
- ✓  $\beta_1$  é o parâmetro de desfasamento do GARCH.  $\beta_1$  mede a persistência da volatilidade condicional.
- ✓ A soma de  $\alpha_1$  e  $\beta_1$  determina quanto tempo permanecerá uma flutuação na projeção. Ou seja, quanto maior a soma, por mais tempo essa flutuação do mercado será incorporada no  $\sigma$  da projeção.
- ✓  $\alpha_0$  é o parâmetro constante e, juntamente com a soma  $\alpha_1$  e  $\beta_1$ , determina o nível da volatilidade média a longo prazo. Esse valor geralmente é muito pequeno para horizontes de dados longos. Também  $\alpha_0$  garante que a volatilidade seja positiva.

Os parâmetros GARCH são estimados pela maximização do valor do estimador de log-verosimilhança (MLE). Ao otimizar um modelo GARCH que tenha um processo de erro após uma distribuição Gaussiana condicional, a função que é maximizada é uma estatística de teste,

$$\ln L(\Theta) = -\frac{1}{2} \sum_{i=1}^I \left( \ln(\sigma_i^2) + \frac{\varepsilon_i^2}{\sigma_i^2} \right) \quad (2.15)$$

onde  $\Theta = \alpha_0, \alpha_1, \beta_1$  são os parâmetros na equação de variância condicional. A função MLE é o conjunto de parâmetros que maximiza a probabilidade de observar os retornos verificados.

Se o modelo for adequado, os resíduos padronizados devem então ser não correlacionados em série.

Outra maneira é olhar para o Ljung-Box-Pierce *Q-Test*.

O *Q-Test* de Ljung-Box-Pierce indica que os resíduos padronizados são realmente não correlacionados em série, o que também é ilustrado na figura 2.2.

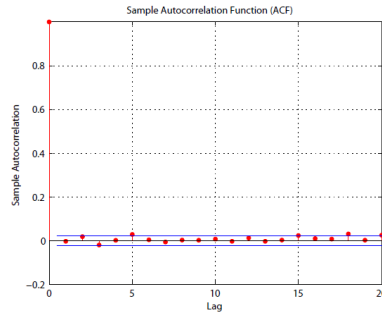


Figura 2.2. Autocorrelação dos resíduos padronizados

Tendo sido estimados os parâmetros do modelo GARCH, torna-se mais fácil estimar o VaR. O  $VaR_\alpha$  pode, para o modelo t-GARCH, ser calculado da seguinte maneira,

$$VaR_{i;\alpha} = \pi_i \cdot \left( \mu + \sigma t_v^{-1} \cdot \sqrt{(v-2)/v(\alpha)} \right)$$

Onde  $t_v$  é a distribuição *t* standard com  $v$  graus de liberdade. Na equação acima, o quantil deve ser dimensionado por um fator  $\sqrt{(v-2)/v}$ .

## 2.2 Métodos Paramétricos e Não-Paramétricos

O VaR é calculado como um quantil da distribuição de ganhos e perdas para um período específico de previsão, tal como já tinha sido referido. Este quantil mostra a pior perda possível, ou seja o maior risco, ao longo do período de tempo escolhido e nível de confiança. Relembrar o exemplo: Calcular o VaR de 95% de 1 dia significa encontrar o quantil de 5% da parte inferior da distribuição de retorno de 1 dia.

Como se sabe todo o cálculo do VaR começa com a configuração de um nível de confiança, escolhendo o período de tempo e estimando a distribuição de probabilidade. Existem, essencialmente, duas maneiras de calcular o VaR: não paramétrico e paramétrico.

Um dos métodos não-paramétrico é conhecido como Simulação Histórica (HS), a estimação do VaR é muito fácil de implementar se os dados ou os retornos estiverem disponíveis. Não há necessidade de quaisquer pressupostos quanto à forma da distribuição dos retornos, apenas que sejam independentes e identicamente distribuídos (i.i.d). Os retornos observados são organizados por tamanho, e dependendo do nível de confiança  $\alpha$  e do número de observações  $n$ , a observação após o primeiro  $\alpha\%$  de retornos é utilizada no cálculo.

$$VaR = -P_0(r^* - \mu), \quad (2.16)$$

onde  $P_0$  é o valor inicial da posição,  $\mu$  é o retorno esperado médio e  $r^*$  é o retorno de *cutoff* para o menor  $\alpha \cdot n$  dos retornos.

Por outro lado, o VaR paramétrico assume que os retornos seguem uma certa distribuição, o que simplifica, significativamente, o cálculo se esta distribuição for da família paramétrica. Por exemplo, para uma distribuição normal dos retornos independentes e identicamente distribuídos (n.i.i.d) precisamos apenas da média  $\mu$  e da volatilidade  $\sigma$  para estimar quantis. Assim, o VaR é calculado da seguinte forma:

$$VaR = P_0 z_\alpha \sigma$$

onde  $z_\alpha$  é o quantil da distribuição para um determinado nível de confiança. Para o VaR paramétrico, como pode ser verificado na fórmula, a precisão da estimativa de volatilidade desempenha um papel fundamental para a figura estimada. E por isso mesmo foram desenvolvidos vários modelos VaR, alguns já indicados anteriormente.

## 2.3 Backtesting

Independentemente do modelo usado para estimar o VaR, ele pode ser comparado com outros modelos, testando-o nos retornos realizados e fornecendo evidências sobre a precisão das previsões.

Normalmente, parte dos dados da amostra é usada para a estimativa (*estimation window*), e as previsões são testadas no restante das observações. Se a perda observada num determinado dia for maior do que a prevista pelo modelo, a violação é registrada. Dependendo do nível de confiança, esperamos violações de 5% e 1% das observações para 95% e 99% de VaR, respetivamente.

No final do *backtest*, o número de violações observadas é comparado ao número de violações previstas. Isso dá a relação de violação para o modelo:

$$\text{Rácio de Violações} = \frac{\text{Número de violações observadas}}{\text{Número de violações previstas}}$$

A proporção deve ser o mais próxima possível, valores acima de 1 significam que o modelo subestima o risco (mais violações do que o previsto acontecer durante o backtesting) e os valores abaixo de 1 são um sinal de modelo “super conservador” (risco superestimado, menos violações do que o previsto). A estimativa e o *backtesting* levam à questão da disponibilidade de dados para previsões do VaR de alto nível de confiança para horizontes temporais superiores a 1 dia. Por exemplo, para 99% de VaR diário, esperamos cerca de 2-3 dias de negociação num ano (1% de 252) com perdas maiores do que as previstas. Isso significa que, para confirmar que o modelo funciona bem, precisamos em média de 20 observações sem violação por 95% de VaR e 100 observações sem violação por 99% de VaR.

Resumindo o *Backtesting* não é mais que um teste estatístico que determina se as estimativas das perdas produzidas pelo modelo de risco estão de acordo com as perdas efetivas (Jorion, 2007).

Nesta secção serão indicados 3 testes diferentes a serem verificados.

1. Mostrar como testar se a **taxa de falha** observada está em linha com o que estava esperado e se existem ou não aglomerados de volatilidade, o que poderia levar a um grupo de violações (*clusters*). Embora isso não mostre diretamente se o conjunto de dados viola o i.i.d.

Uma vez que o modelo de Basileia II assume retornos para seguir uma distribuição gaussiana, isso implica indiretamente uma hipótese de que os retornos são i.i.d. Como essas premissas são fundamentais para o próprio modelo, é de grande importância testar se elas são ou não válidas. Para isso recorre-se a dois testes:

2. O teste **Ljung-Box** é um teste de aleatoriedade de observações que, portanto, testa se os dados são distribuídos de forma independente. A falha neste teste significa uma clara violação da suposição i.i.d.
3. O teste **Jarque-Bera** permite também testar a violação do pressuposto i.i.d

Estes três testes mostram quão bem os pressupostos no modelo de Basileia II se enquadram nas séries empíricas de retorno.

### 2.3.1 Teste de taxa de falha

Explicitamente, as taxas de falha observadas fornecem uma medida quantitativa da precisão do modelo. Intuitivamente, uma taxa de falha de 1% seria esperada para VaR. e VaR+. O problema, no entanto, é que é necessário um teste estatístico para ver se a taxa de falha real percebida é muito alta ou muito baixa em comparação com a taxa de falha esperada. É importante entender que ambos os resultados resultarão em um modelo defeituoso. Isto ocorre se o número de violações for muito baixo, o modelo é muito conservador, levando a uma alocação de capital ineficaz. Daí que a empresa que aplica o modelo não atuará nos melhores interesses dos seus proprietários (acionistas, etc.). É de igual importância notar que, se o modelo tivesse 75 violações do VaR, este ainda pode estar com defeito.

Por exemplo, as violações podem vir em grandes grupos de dados (*clusters*), enquanto que, noutras ocasiões, o modelo pode superestimar drasticamente o risco como já foi observado. Portanto, é preciso ter cuidado não só por rejeitar um modelo correto, mas também para não aceitar um modelo defeituoso.

### 2.3.2 Teste de Kupiec e Christofferson

Indiretamente, o teste Ljung-Box testa a correlação entre os retornos. Se não houver correlação entre os retornos, então o melhor será testar a média da distribuição. No entanto, se houver evidência de correlação, isso contradiz a hipótese de passeio aleatório e a melhor estimativa para o retorno futuro pode ser feita com um modelo condicional. Isto significa que se não houver provas de correlação significativas, os retornos não podem ser assumidos de forma independente a partir da mesma distribuição. Consequentemente, o teste de Ljung-Box é um teste da hipótese geral da suposição i.i.d de retornos, onde o  $LR_m$  apenas testou a independência entre violações.

As estatísticas de teste são dadas por,

$$Q_{LB} = n(n+2) \sum_{j=1}^h \frac{\hat{p}(j)^2}{n-j} \quad (2.17)$$

onde o  $\hat{p}(j)$  é o fator de correlação entre a observação  $jth$  e a observação inicial. O teste de Ljung-Box é, portanto, um teste da previsibilidade entre as observações. Traçar o  $\hat{p}(j)$  contra  $j$  dá uma função de autocorrelação (ACF). Este gráfico fornece uma ilustração facilmente interpretada que é útil para detetar quaisquer correlações dentro da série de dados.

Como um modelo é apenas uma aproximação da realidade, uma taxa de falha em, precisamente, 1% provavelmente não será observado. Além disso, podem haver outros pressupostos que podem ser encontrados e não se mantêm quando aplicados em dados empíricos.

Apesar de se observar uma taxa de falha em condições normais de mercado próximas de 1% tanto para VaR<sub>-</sub> como para VaR<sub>+</sub> num horizonte temporal de um dia, houve indícios de que o modelo de Basileia II não conseguiu capturar a heterocedasticidade da volatilidade, bem como o comportamento leptocúrtico do retorno distribuição. Isso também pode violar o modelo e, portanto, deve ser testado assim como a taxa de falha.

Testando a taxa de falha, sabe-se que há dois resultados possíveis em cada ponto no tempo, ou seja, o VaR pode ser violado ou não. Por isso, cada resultado diário pode ser tratado como um teste de Bernoulli com uma distribuição binomial.

Considerando  $n$  como o número total de ensaios,  $p$  a probabilidade de sucesso assumida (a probabilidade de violação do VaR) e  $x$  o número de sucessos na série então a equação abaixo pode ser usada para estimar a taxa de falha.

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ onde } x = 0, 1, 2, \dots, n \quad (2.18)$$

Isto porque é de todo o interesse testar se o modelo deve ou não ser rejeitado, as proporções opostas devem ser levantadas, ou seja, se,

$$H_0: p = p_0 = 1\%$$

Como o teste de proporções tem os dois lados (ser rejeitado ou não), ele indica o *spread* dentro do qual a taxa de falha da amostra estará em linha com a taxa de falha da população. Se essa hipótese for rejeitada, o modelo deve ser rejeitado. Observe-se que este teste deve ser realizado para VaR<sub>-</sub> e VaR<sub>+</sub>. No entanto, devido à simetria da distribuição, o *spread* será o mesmo.

A escolha do intervalo de confiança do teste é um equilíbrio entre a precisão do teste e o intervalo em que os valores de teste são encontrados. Por exemplo, se for escolhido um intervalo de confiança de 100%, o teste mostraria que, com 100% de probabilidade, a taxa de falha ficaria entre 0 e  $\infty$ . Diminuir o intervalo de confiança também diminuirá o intervalo em que a taxa de falha deveria ficar para que a hipótese seja aceite. Geralmente, um intervalo de confiança de 95% é o mais indicado nestes testes de hipóteses. Este intervalo de equivalência é relativamente preciso e ao mesmo tempo dá um intervalo suficientemente pequeno. Esta é a razão pela qual é aplicado.

Um teste mais científico que permitirá aceitar ou rejeitar o modelo com base na taxa de falhas, é um teste da verossimilhança<sup>(1)</sup>. Este método também testa se as violações são ou não distribuídas independentemente e, portanto, se a distribuição das violações está em conformidade com a suposição i.i.d.

Propondo a hipótese nula:

$$H_0: p = p_0 = 1\%$$

contra a hipótese alternativa:

$$H_1: p \neq p_0$$

<sup>(1)</sup> Usa um estimador de verossimilhança para testar a cobertura incondicional e a independência do modelo e combina estes dois numa única medida para a cobertura condicional, que descreve a precisão geral do modelo.

combinada com a equação acima indicada, que afirma que a probabilidade de ver  $n_1$  violações é,

$$(1 - p_0)^{n_0} p_0^{n_1}$$

onde  $n_0 = n - n_1$  é o número total das não-violações observadas, é possível testar a probabilidade dessa hipótese em relação à probabilidade observada de  $n_1$ ,

$$(1 - \pi)^{n_0} \pi^{n_1}$$

onde  $\pi$  é a estimativa da taxa de falha real,  $\pi \approx \hat{\pi} = n_1/n_0$ . Isto dá o teste log-verossimilhança indicado abaixo,

$$\begin{aligned} LR_{uc} &= -2 \cdot \log \left[ \frac{(1 - p_0)^{n_0} p_0^{n_1}}{(1 - \pi)^{n_0} \pi^{n_1}} \right] \sim \chi^2_{1-\alpha}(1) \\ &= -2 \cdot [n_0 \cdot \log(1 - p_0) + n_1 \cdot \log(p_0) - n_0 \cdot \log(1 - \hat{\pi}) - n_1 \cdot \log(\hat{\pi})] \end{aligned} \quad (2.19)$$

onde  $\chi^2_{1-\alpha}(1)$  é o qui-quadrado de  $1-\alpha$  com 1 grau de liberdade.

Este é um teste de cobertura incondicional e explica o poder de ajustamento da taxa de falha em comparação com a taxa de falha proposta em  $H_0$ .  $H_0$  não é rejeitado quando  $LR_{uc} < \chi^2_{1-\alpha}(1)$ , o que significa que a taxa de falha geral observada está em linha com a taxa de falha esperada. Se este não for o caso,  $H_0$  será rejeitado e  $H_1$  será aceite.

Este teste de cobertura incondicional não prova o *clustering*<sup>(2)</sup> e, portanto, não revela se existe uma tendência para grandes violações entrar em *clusters*. Isso, no entanto, pode ser feito com uma prova de independência.

Quando as violações não são independentes, a probabilidade de uma violação amanhã, dado que houve uma violação hoje, não é mais igual a  $p$ . Por isso, é necessário configurar um teste que revele essas tendências.

Assim define-se,

- ✓  $n_{01}$  corresponde ao número de observações em que uma não violação é seguida por uma violação
- ✓  $n_{11}$  quando uma violação é seguida por outra violação
- ✓  $n_{10}$  quando uma violação é seguida por uma não-violação
- ✓  $n_{00}$  quando uma não-violação é seguida por outra não-violação

$\Pi_1$  pode ser definido da seguinte maneira,

$$\begin{aligned} \Pi_1 &= \begin{pmatrix} \pi_{01} \\ \pi_{11} \end{pmatrix} \\ &= \begin{pmatrix} \frac{n_{01}}{n_{00} + n_{01}} \\ \frac{n_{11}}{n_{10} + n_{11}} \end{pmatrix} \end{aligned}$$

O teste do rácio log-verossimilhança para independência torna-se assim um teste para a hipótese nula

$$H_0: \Pi_1 = \pi_2$$

<sup>(2)</sup> Clustering é o método de separar os seus dados em grupos (clusters) quando estamos a agrupar dados, tendo por base as semelhanças.



contra a hipótese

$$H_1: \Pi_1 \neq \pi_2$$

com as estatísticas,

$$LR_{in} = -2 \cdot \log \frac{(1 - \pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1 - \pi_{11})^{n_{10}} \pi_{11}^{n_{11}}}{(1 - \hat{\pi}_2)^{(n_{00} + n_{10})} \hat{\pi}_2^{(n_{01} + n_{11})}} \sim \chi^2_{1-\alpha}(1) \quad (2.20)$$

onde  $\pi_2 \approx \hat{\pi}_2 = (n_{01} + n_{11})/(n)$ .

O que esta hipótese faz é testar se a taxa de falha geral é a mesma que a probabilidade de se observar uma violação após uma violação. Se isto for verdade, a série de violações pode ser considerada independente e, portanto, pode-se assumir que as violações não se agrupam.

Este teste ignora completamente o  $p_0$  e apenas testa se é provável que a probabilidade de  $n_{11}$  e  $n_{01}$  seja a mesma que a probabilidade geral de observar uma violação.

Se esse teste testar que o valor  $p$  é muito alto em comparação com o valor do qui-quadrado, o teste de independência é rejeitado, com a consequência de que as violações não são i.i.d.

É de destacar que se  $LR_{uc}$  é rejeitado, mas  $LR_{in}$  não é rejeitado, então temos um caso especial, onde o  $p \neq p_0$ , mas onde as violações são i.i.d. Por outro lado, se o teste de cobertura incondicional e independência não for rejeitado, isto significa que a taxa de falha observada é próxima do nível de confiança e que a probabilidade de observar as violações subsequentes também é próxima desse mesmo nível de confiança.

## 2.4 Extreme Value Theory (EVT)

Nesta seção vai ser descrita uma breve apresentação da teoria dos valores extremos (EVT) e de que maneira é que essa teoria é particularmente útil para modelar o VaR no mercado de condições extremas.

A maioria dos métodos estatísticos concentram-se em toda a distribuição. Ao contrário da EVT, em que o seu foco é apenas nas caudas da distribuição - eventos extremamente raros.

Recentemente, inúmeros estudos de pesquisa analisaram as variações extremas às quais os mercados financeiros estão sujeitos, principalmente devido a crises cambiais, falhas no mercado de ações e grandes incumprimentos de crédito. O comportamento da cauda das séries financeiras foi discutido por Koedijk *et al.* (1990), Dacorogna *et al.* (1995), Loretan e Phillips (1994), McNeil e Frey (2000), entre outros. Uma discussão interessante sobre o potencial da teoria do valor extremo na gestão de risco é dada por Diebold *et al.* (1998).

A EVT é especialmente útil para o cálculo do VaR com um alto nível de confiança, o que reflete as perdas extremamente raras na cauda de distribuição. Os *inputs* chave para o modelo EVT são o índice da cauda e o inverso dos parâmetros do índice da cauda.

O VaR é geralmente definido como o capital suficiente para cobrir, na maioria dos casos, as perdas de uma carteira ao longo de um horizonte temporal. Supondo que uma variável aleatória  $X$  com função de distribuição contínua  $F$  modela perdas ou retornos negativos num determinado instrumento financeiro ao longo de um determinado horizonte temporal, o  $VaR_p$  pode então ser definido como a  $p$  quantil da distribuição  $F$ ,

$$VaR_p = F^{-1}(1 - p) \quad (2.21)$$

onde  $F^{-1}$  é designado como o quantil da função, sendo este a inversa da distribuição da função  $F$ .

Para fins de controlo de risco interno, a maioria das empresas calcula um VaR de 5% ao longo de um período temporal de um dia. Recorde-se que o acordo de Basileia propôs que para o cálculo do VaR seja considerado um horizonte temporal de 10 dias e  $p = 1\%$ , com base num período de observação histórico de pelo menos 1 ano de dados diários, deve ser calculado e depois multiplicado pelo "*safety factor*". Este fator de segurança foi introduzido porque a hipótese normal para a distribuição de lucros e perdas é amplamente reconhecida como não sendo realista.

Ao modelar os máximos de uma variável aleatória, EVT desempenha o mesmo papel fundamental que o teorema do limite central ao modelar somas de variáveis aleatórias. Em ambos os casos, a teoria diz-nos quais são as distribuições limitantes.

Geralmente, existem duas maneiras relacionadas de identificar os extremos em dados reais. Considere-se uma variável aleatória que representa as perdas ou os retornos diários. A primeira abordagem considera o máximo que a variável leva em períodos sucessivos, por exemplo, meses ou anos. Essas observações selecionadas constituem os eventos extremos, também chamados de *block maxima*.

O *block maxima method* é o método tradicional usado para analisar dados com sazonalidade. No entanto, o *threshold method* usa os dados de forma mais eficiente e, por essa razão, parece tornar-se o método mais usado para aplicações recentes.

No seguimento da tese irão ser descritos estes dois métodos, o *block maxima method* e o *threshold method*.

### 2.4.1 Distribuição dos valores máximos

A lei limite para os blocos de máximo, que denota-se por  $M_n$ , sendo  $n$  o tamanho da subamostra (bloco), é dada pelo seguinte teorema:

**Teorema 1** (Fisher e Tippet (1928), Gnedenko (1943))

Seja  $(X_n)$  uma sequência de variáveis aleatórias i.i.d. Se houver constantes  $c_n > 0$ ,  $d_n \in \mathbb{R}$  e alguma função de distribuição não degenerada  $H$  tal que,

$$\frac{M_n - d_n}{c_n} \xrightarrow{d} H \quad (2.22)$$

Então  $H$  pertence a uma das três distribuições de valores extremos padrão:

Fréchet:

$$\phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ e^{-x^{-\alpha}}, & x > 0 \end{cases} \quad \alpha > 0 \quad (2.23)$$

Weibull:

$$\psi_\alpha(x) = \begin{cases} e^{-(-x)^\alpha}, & x \leq 0 \\ 1, & x > 0 \end{cases} \quad \alpha > 0 \quad (2.24)$$

Gumbel:

$$\Lambda(x) = e^{-e^{-x}}, x \in \mathbb{R} \quad (2.25)$$

A forma das funções de densidade de probabilidade para as distribuições padrão de Fréchet, Weibull e Gumbel é dada pela figura 2.3, respetivamente.

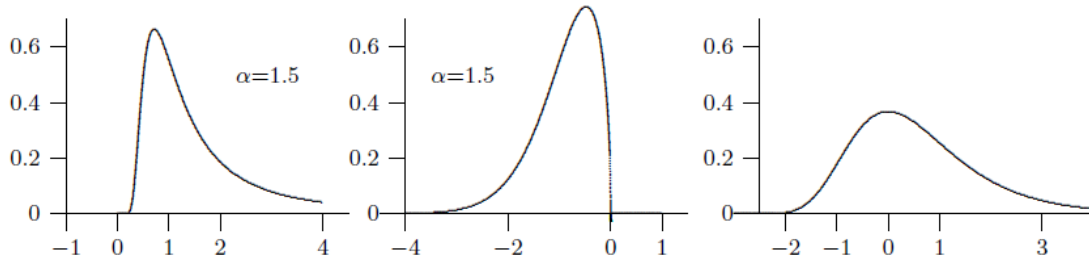


Figura 2.3. Funções de densidade de Fréchet, Weibull e Gumbel (fonte Manfred Gilli e Evis KÄellezi (2006))

Com a análise da figura 2.3 pode-se observar que a distribuição de Fréchet tem uma cauda de decomposição polinomial e, portanto, combina as distribuições de caudas bem pesadas. As caudas exponencialmente decedentes da distribuição de Gumbel caracterizam distribuições de cauda fina. Por fim, a distribuição de Weibull é a distribuição assintótica das distribuições finais de pontos finais.

Jenkinson (1955) e von Mises (1954) sugeriram a seguinte representação de um parâmetro:

$$H_{\xi}(x) = \begin{cases} e^{-(1+\xi x)^{-1/\xi}} & \text{se } \xi \neq 0 \\ e^{-e^{-x}} & \text{se } \xi = 0 \end{cases} \quad (2.26)$$

dessas três distribuições padrão, com  $x$  tal que  $1 + \xi x > 0$ . Esta generalização, conhecida como distribuição de *generalized extreme value* (GEV), é obtida configurando  $\xi = \alpha^{-1}$  para a distribuição de Fréchet,  $\xi = -\alpha^{-1}$  para a distribuição de Weibull e interpretando a distribuição de Gumbel como o caso limite  $\xi = 0$ .

Antecipadamente, não se sabe o tipo de distribuição restante dos máximos da amostra, a representação generalizada é particularmente útil quando as estimativas de máxima verossimilhança devem ser computadas. Além disso, o padrão GEV definido no sistema acima é a distribuição limitante de extrema normalizada. Dado que, na prática, não se conhece a verdadeira distribuição dos retornos e, como resultado, não se tem nenhuma ideia sobre as constantes de normalização  $c_n$  e  $d_n$ , usa-se a especificação de três parâmetros da GEV, que é a distribuição limitante dos máximos não normalizados,

$$H_{\xi, \sigma, \mu}(x) = H_{\xi}\left(\frac{x - \mu}{\sigma}\right) \quad x \in D, \quad D = \begin{cases} ] - \infty, \mu - \frac{\sigma}{\xi} [ & \xi < 0 \\ ] - \infty, \infty [ & \xi = 0 \\ ] \mu - \frac{\sigma}{\xi}, \infty [ & \xi > 0 \end{cases} \quad (2.27)$$

Os dois parâmetros adicionais  $\mu$  e  $\sigma$  são a localização e os parâmetros da escala que representam as constantes de normalização desconhecidas.

As quantidades de interesse não são os próprios parâmetros, mas os quantis, também chamados níveis de retorno, da GEV estimada:

$$R^k = H_{\xi, \sigma, \mu}^{-1}\left(1 - \frac{1}{k}\right) \quad (2.28)$$

Substituindo os parâmetros  $\xi, \sigma$  e  $\mu$  pelas suas estimativas  $\hat{\xi}, \hat{\sigma}$  e  $\hat{\mu}$  tem-se,

$$\hat{R}^k = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} \left( 1 - \left( -\log \left( 1 - \frac{1}{k} \right) \right)^{-\hat{\xi}} \right) & \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma} \log \left( -\log \left( 1 - \frac{1}{k} \right) \right) & \hat{\xi} = 0 \end{cases} \quad (2.29)$$

Um valor de  $\hat{R}^{10}$  de 5 significa que, em média, a perda máxima observada durante um período de um ano será superior a 5% uma vez em dez anos.

A aplicação do método dos blocos de máximos passa pelas seguintes etapas: dividir a amostra em  $n$  blocos de igual comprimento, seleccionar o valor máximo em cada bloco, ajustar a distribuição de GEV ao conjunto de máximos e, finalmente, estimar o ponto de cálculo e intervalo para  $R_n^k$ .

O ponto delicado deste método é a escolha apropriada dos períodos que definem os blocos. O calendário naturalmente sugere períodos como meses, trimestres, etc. Para evitar efeitos sazonais, escolhe-se períodos anuais que provavelmente serão suficientemente grandes para que o Teorema 1 seja mantido. Na maioria dos casos nem todos os blocos são exatamente do mesmo comprimento. O retorno máximo em cada um dos blocos constitui os pontos de dados para a amostra de máximos  $M$  que é utilizada para estimar a distribuição de valor extremo generalizada (GEV).

A função log-verossimilhança que se maximiza em relação aos três parâmetros desconhecidos é,

$$L(\xi, \mu, \sigma; x) = \sum_i \log(h(x_i)), \quad x_i \in M \quad (2.30)$$

onde,

$$h(\xi, \mu, \sigma; x) = \frac{1}{\sigma} \left( 1 + \xi \frac{x - \mu}{\sigma} \right) \exp \left( -\exp \left( -\frac{x - \mu}{\sigma} \right) \right)$$

Para se poder calcular as estimativas do intervalo, é útil abordar o problema da estimativa do quantil reparametrizando diretamente a distribuição GEV em função do nível de retorno desconhecido  $R_k$ . Para se conseguir isso, isola-se  $\mu$  da seguinte equação,

$$\hat{R}^k = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} \left( 1 - \left( -\log \left( 1 - \frac{1}{k} \right) \right)^{-\hat{\xi}} \right) & \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma} \log \left( -\log \left( 1 - \frac{1}{k} \right) \right) & \hat{\xi} = 0 \end{cases} \quad (2.31)$$

e substituí-lo em  $H_{\xi, \sigma, \mu}$  definido na equação (2.27), a função de distribuição GEV resulta então,

$$H_{\xi, \sigma, R^k}(x) = \begin{cases} \exp \left( - \left( \frac{\xi}{\sigma} (x - R^k) + \left( -\log \left( 1 - \frac{1}{k} \right) \right)^{-\xi} \right)^{-\frac{1}{\xi}} \right) & \xi \neq 0 \\ \left( 1 - \frac{1}{k} \right)^{\exp \left( -\frac{x - R^k}{\sigma} \right)} & \xi = 0 \end{cases} \quad (2.32)$$

para  $x \in D$  definido como,

$$D = \begin{cases} ] - \infty, \left(R^k - \frac{\xi}{\sigma} \left(-\log\left(1 - \frac{1}{k}\right)\right)^{-\xi}\right) [ & \xi < 0 \\ ] - \infty, \infty [ & \xi = 0 \\ ] \left(R^k - \frac{\xi}{\sigma} \left(-\log\left(1 - \frac{1}{k}\right)\right)^{-\xi}\right), \infty [ & \xi > 0 \end{cases} \quad (2.33)$$

Pode-se obter diretamente estimativas de máxima verossimilhança para  $R^k$ . A função de log-verossimilhança de perfil pode então ser usada para calcular intervalos de confiança separados ou conjuntos para cada um dos parâmetros. Por exemplo, no caso em que o parâmetro de interesse é  $R^k$ , a função de log-verossimilhança de perfil será definida como,

$$L^*(R^k) = \max_{\xi, \sigma} L(\xi, \sigma, R^k). \quad (2.34)$$

O intervalo de confiança que foi derivado inclui, assim, todos os valores de  $R^k$  que satisfazem a condição,

$$L^*(R^k) - L(\hat{\xi}, \hat{\sigma}, \hat{R}^k) > -\frac{1}{2} \chi_{\alpha,1}^2 \quad (2.35)$$

onde  $\chi_{\alpha,1}^2$  refere-se ao  $(1 - \alpha)$  quantil de nível da distribuição  $\chi^2$  com 1 grau de liberdade.

A função  $L^*(R^k) - L(\hat{\xi}, \hat{\sigma}, \hat{R}^k)$  é chamada de *relative profile log-likelihood function*.

Por vezes, também se tem interesse no valor de  $\xi$ , que caracteriza o peso da cauda da distribuição subjacente. A função *profile log-likelihood* é então definida como,

$$L^*(\xi, R^k) = \max_{\sigma} L(\xi, \sigma, R^k), \quad (2.36)$$

Com a região de confiança definida como o contorno no nível  $-\frac{1}{2} \chi_{\alpha,2}^2$  da função *relative profile log-likelihood*

$$L^*(\xi, R^k) - L(\hat{\xi}, \hat{\sigma}, \hat{R}^k). \quad (2.37)$$

### 2.4.2 The Peak Over Threshold Method

Uma abordagem alternativa é o método conhecido como *peak over threshold* (POT), este método considera a distribuição de excedências num certo limiar. O problema está ilustrado na figura abaixo onde se considera uma função de distribuição (desconhecida)  $F$  de uma variável aleatória  $X$ . Tem-se como objetivo estimar a função de distribuição  $F_u$  de valores de  $x$  acima de um determinado limite  $u$ .

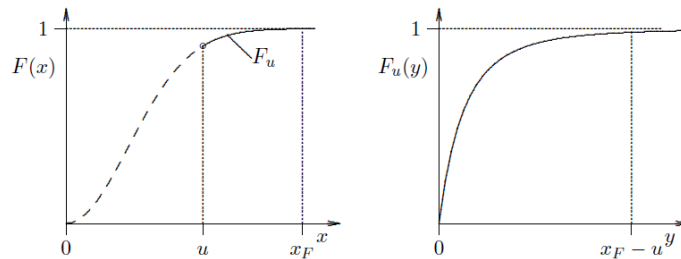


Figura 2.4. Distribuição da função  $F$  e distribuição condicional da função  $F_u$  (fonte Manfred Gilli e Evis KÄellezi (2006))

A distribuição da função  $F_u$  é chamada de *conditional excess distribution function* e é definida como,

$$F_u(y) = P(X - u \leq y | X > u), \quad 0 \leq y \leq x_F - u \quad (2.38)$$

onde  $X$  é uma variável aleatória,  $u$  é um determinado limite,  $y = x - u$  são os excessos e  $x_F \leq \infty$  é o ponto final direito de  $F$ , isto é,

$$F_u(y) = \frac{F(u + y) - F(u)}{1 - F(u)} = \frac{F(x) - F(u)}{1 - F(u)} \quad (2.39)$$

As realizações da variável aleatória  $X$  situam-se principalmente entre 0 e  $u$  e, portanto, a estimativa de  $F$  neste intervalo geralmente não apresenta problemas. A estimativa de  $F_u$ , no entanto, pode ser difícil, pois geralmente, observa-se muito poucas observações nessa área.

Neste ponto, a EVT pode ser muito útil, uma vez que proporciona um forte resultado sobre a *conditional excess distribution function* que está estabelecida no seguinte teorema:

**Teorema 2:** (Pickands (1975), Balkema e de Haan (1974))

Para uma grande classe de funções de distribuição subjacentes  $F$ , a função de distribuição de excesso condicional  $F_u(y)$ , para  $u$ , é bem aproximada por,

$$F_u(y) \approx G_{\xi, \sigma}(y), \quad u \rightarrow \infty$$

onde,

$$G_{\xi, \sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma}y\right)^{-\frac{1}{\xi}} & \text{se } \xi \neq 0 \\ 1 - e^{-\frac{y}{\sigma}} & \text{se } \xi = 0 \end{cases} \quad (2.40)$$

para  $y \in [0, (x_F - u)]$  se  $\xi \geq 0$  e  $y \in \left[0, -\frac{y}{\sigma}\right]$  se  $\xi < 0$ .  $G_{\xi, \sigma}$  é chamada a distribuição de Pareto generalizada (GPD).

Se  $x$  é definido como  $x = u + y$ , o GPD também pode ser expresso como uma função de  $x$ , isto é,

$$G_{\xi, \sigma}(x) = 1 - (1 + \xi(x - u)/\sigma)^{-1/\xi} \quad (2.41)$$

A figura 2.5 ilustra a forma da distribuição generalizada de Pareto  $G_{\xi, \sigma}(x)$  quando  $\xi$ , chamado *shape parameter* ou *tail index*, tem um valor negativo, positivo e zero. O parâmetro de escala  $\sigma$  é mantido igual a um.

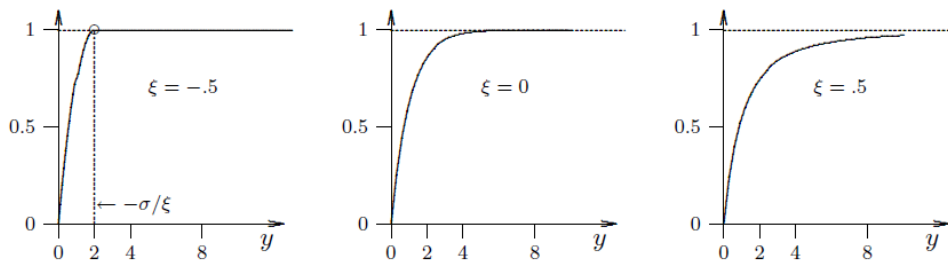


Figura 2.5. Distribuição generalizada de Pareto  $G_{\xi, \sigma}(x)$  para  $\sigma=1$  (fonte Manfred Gilli e Evis KÄellezi (2006))

A *tail index*  $\xi$  dá uma indicação do peso da cauda, quanto maior  $\xi$ , mais pesada será a cauda. Como, em geral, não é possível consertar um limite superior para perdas financeiras, apenas distribuições com *shape parameter*  $\xi \geq 0$ , são adequadas para modelar distribuições de retorno financeiro.

Assumindo a função GPD para a distribuição da cauda, as expressões analíticas para  $Var_p$  e  $ES_p$  podem ser definidas como os parâmetros de uma função GPD. Isolando  $F(x)$  da equação  $F_u(y)$ ,

$$F(x) = (1 - F(u)) F_u(y) + F(u) \quad (2.42)$$

e substituindo  $F_u$  pela GPD e  $F(u)$  pela estimativa  $(n - N_u)/n$ , onde  $n$  é o número total de observações e  $N_u$  o número de observações acima do limite  $u$ , obtém-se

$$\hat{F}(x) = \frac{N_u}{n} \left( 1 - \left( 1 + \frac{\xi}{\sigma} (x - u) \right)^{-1/\xi} \right) + \left( 1 - \frac{N_u}{n} \right) \quad (2.43)$$

que simplifica para,

$$\hat{F}(x) = 1 - \frac{N_u}{n} \left( 1 + \frac{\xi}{\sigma} (x - u) \right)^{-1/\xi} \quad (2.44)$$

Invertendo a equação acima (2.44) para uma dada probabilidade  $p$  resulta,

$$\widehat{Var}_p = u + \frac{\xi}{\sigma} \left( \left( \frac{n}{N_u} p \right)^{-\xi} - 1 \right) \quad (2.45)$$

Reescreve-se a *expected shortfall* como,

$$\widehat{ES}_p = \widehat{Var}_p + E(X - \widehat{Var}_p | X > \widehat{Var}_p)$$

onde o segundo termo à direita é o valor esperado das excedências sobre o limite  $Var_p$ .

Sabe-se que a *mean excess function* para o GPD com o parâmetro  $\xi < 1$  é

$$e(z) = E(X - z | X > z) = \frac{\sigma + \xi z}{1 - \xi}, \quad \sigma + \xi z > 0 \quad (2.46)$$

Esta função dá a média dos excessos de  $X$  para os valores variáveis de um limiar  $z$ . Outro resultado importante em relação à existência de momentos é que, se o  $X$  seguir um GPD, então, para todos os inteiros  $r$ , tais como  $r < 1/\xi$ , os primeiros momentos de  $r$  existem.

Da mesma forma, dada a definição para o *expected shortfall* e usando a equação acima, para  $z = Var_p - u$  e  $X$  representando os excessos acima obtém-se,

$$\widehat{ES}_p = \widehat{Var}_p + \frac{\hat{\sigma} + \xi(\widehat{Var}_p - u)}{1 - \xi} = \frac{\widehat{Var}_p}{1 - \xi} + \frac{\hat{\sigma} + \xi u}{1 - \xi}. \quad (2.47)$$

A implementação do método de *peak over threshold* envolve as seguintes etapas: selecionar o limite  $u$ , ajustar a função GPD para as excedências sobre  $u$  e, em seguida, calcular as estimativas de ponto e intervalo para *Value-at-Risk* (VaR) e o *expected shortfall*.

Comece-se por referir como selecionar o limite  $u$ .

A teoria refere que  $u$  deve ser alto para satisfazer o Teorema 2, mas quanto maior o *threshold*, menos observações são deixadas para a estimativa dos parâmetros da função de distribuição da cauda.

Até agora, nenhum algoritmo automático com desempenho satisfatório para a seleção do *threshold*  $u$  está disponível. A questão da determinação da fração de dados pertencentes à cauda é tratada por Danielsson *et al.* (2001), Danielsson e de Vries (1997) e Dupuis (1998) entre outros. No entanto, essas referências não fornecem uma resposta clara à questão de qual método deve ser usado.

Uma ferramenta gráfica que é muito útil para a seleção do *threshold*  $u$  é a *sample mean excess plot* definida pelos pontos

$$(u, e_n(u)), \quad x_1^n < u < x_n^n,$$

onde  $e_n(u)$  é a *sample mean excess function* definida como,

$$e_n(u) = \frac{\sum_{i=k}^n (x_i^n - u)}{n - k + 1}, \quad k = \min\{i | x_i^n > u\}, \quad (2.48)$$

e  $(n - k + 1)$  é o número de observações excedendo o *threshold*  $u$ .

A *sample mean excess function* (2.46), que é uma estimativa da função excedente média  $e(u)$  deve ser linear. Esta propriedade pode ser usada como critério para a seleção de  $u$ .

Para a estimação da máxima verossimilhança, sabe-se que a distribuição das observações acima do *threshold* na cauda deve ser uma distribuição generalizada de Pareto (GPD). Existem vários métodos diferentes que podem ser usados para estimar os parâmetros do GPD. Nesta subseção, será descrito o método de estimação da máxima verossimilhança.

Para uma amostra  $y = \{y_1, \dots, y_n\}$  a função log-verossimilhança  $L(\xi, \sigma | y)$  para o GPD é o logaritmo da densidade conjunta das  $n$  observações,

$$L(\xi, \sigma | y) = \begin{cases} -n \log \sigma - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^n \log \left(1 + \frac{\xi}{\sigma} y_i\right) & \text{se } \xi \neq 0 \\ -n \log \sigma - \frac{1}{\sigma} \sum_{i=1}^n y_i & \text{se } \xi = 0 \end{cases} \quad (2.49)$$

Calcula-se os valores  $\hat{\xi}$  e  $\hat{\sigma}$  que maximizam a função de verossimilhança para a amostra decidida pelas observações que excedem o *threshold*  $u$ .

Para as estimativas de intervalo, mais uma vez, considerou-se os intervalos de confiança única e conjunta, com base nas funções de log-verossimilhança do perfil. Os intervalos de confiança baseados em log-verossimilhança para  $VaR_p$  podem ser obtidos usando uma versão reparada de GPD definida como uma função de  $\xi$  e  $VaR_p$ :



$$G_{\xi, VaR_p}(y) = \begin{cases} 1 - \left( 1 + \frac{\left(\frac{n}{N_u}p\right)^{-\xi} - 1}{VaR_p - u} y \right)^{-\frac{1}{\xi}} & \xi \neq 0 \\ 1 - \frac{n}{N_u} p \exp\left(\frac{y}{VaR_p - u}\right) & \xi = 0 \end{cases} \quad (2.50)$$

A função de densidade de probabilidade correspondente é,

$$g_{\xi, VaR_p}(y) = \begin{cases} \frac{\left(\frac{n}{N_u}p\right)^{-\xi} - 1}{\xi(VaR_p - u)} \left( 1 + \frac{\left(\frac{n}{N_u}p\right)^{-\xi} - 1}{VaR_p - u} y \right)^{-\frac{1}{\xi}-1} & \xi \neq 0 \\ -\frac{\frac{n}{N_u} p \exp\left(\frac{y}{VaR_p - u}\right)}{VaR_p - u} & \xi = 0 \end{cases} \quad (2.51)$$

Da mesma forma, usando a seguinte reparametrização para  $\xi \neq 0$

$$G_{\xi, ES_p} = 1 - \left( 1 + \frac{\xi + \left(\frac{n}{N_u}p\right)^{-\xi} - 1}{(ES_p - u)(1 - \xi)} y \right)^{-\frac{1}{\xi}},$$

$$g_{\xi, ES_p} = \frac{\xi + \left(\frac{n}{N_u}p\right)^{-\xi} - 1}{\xi(1 - \xi)(ES_p - u)} \left( 1 + \frac{\xi + \left(\frac{n}{N_u}p\right)^{-\xi} - 1}{(ES_p - u)(1 - \xi)} y \right)^{-\frac{1}{\xi}-1}, \quad (2.52)$$

Por fim calcula-se um intervalo de conflito baseado na log-verossimilhança para o *expected shortfall*  $ES_p$ .



## Capítulo III

### Modelos Estimados/Estimação e Validação do VaR

Neste capítulo irá ser feita a análise dos dados considerados (PSI-20) e a implementação dos modelos e testes econométricos relacionados com a previsão do VaR, tendo em conta os conceitos mencionados no capítulo anterior.

A série utilizada nesta dissertação tem 2892 observações diárias, 5 dias por semana, no período temporal compreendido entre 2 de Janeiro de 2006 até 31 de Janeiro de 2017. Os dados da série foram retirados da base de dados DATASTREAM e refletem os preços de fecho do índice PSI-20. Como é usual em Economia e Finanças, as séries não são trabalhadas em níveis, mas sim em taxas de crescimento e rendibilidades (retornos logaritmizados), respetivamente.

A série irá ser analisada quanto à sua estacionaridade e normalidade. A suposição de normalidade é crucial para alguns métodos de cálculo do VaR. Isso é bastante semelhante a muitos outros modelos financeiros, uma vez que a normalidade simplifica muito os cálculos de dependências que, de outra forma, podem tornar-se complexas. Uma vez que a distribuição normal pode ser descrita apenas pela média e pelo desvio padrão, o cálculo do VaR é o mais simples quando a suposição de normalidade se mantém. Mas a metodologia não se separa quando os retornos não são normalmente distribuídos, embora o VaR possa perder algumas propriedades desejáveis (Artzner *et al.*, 1999).

#### 3.1 Análise descritiva e testes de raiz unitária

Inicialmente na análise da série considerada, estuda-se a representação gráfica da série temporal em níveis (figura 3.1) e dos retornos logaritmizados (figura 3.2).

Pela observação do gráfico da figura 3.1, verifica-se que evidencia um comportamento irregular, sem uma tendência monótona estabelecida. Verifica-se, ainda, que os picos foram nos anos 2006/2007, 2009 a 2011 e 2013/2014. Este resultado de certo modo tem sentido, tendo em conta que a crise financeira em Portugal foi nos anos de 2010 a 2014. Adiantamos que a série PSI-20 em níveis aparenta ser não estacionária.

Enquanto que, na figura 3.2, verifica-se que apesar de ter uma média constante, ainda apresenta valores extremos nas datas entre 2007-2009, em 2010 e entre 2015-2016. Nesta série de retornos conseguimos verificar a existência de *clusters* (grupos de intervalos repetitivos de valores muito pequenos ou muito grandes) de volatilidade.

A partir da análise das suas estatísticas descritivas verifica-se que o *P-value* associado ao teste de normalidade de Jarque-Bera é aproximadamente 0.000 e, portanto, as séries não seguem uma distribuição normal. Pois, considerando a hipótese nula do teste de Jarque-Bera:  $H_0$ : Distribuição Normal e um intervalo de confiança de 95%, verifica-se que  $0.000 < 0.05$  rejeitando-se assim a hipótese nula.

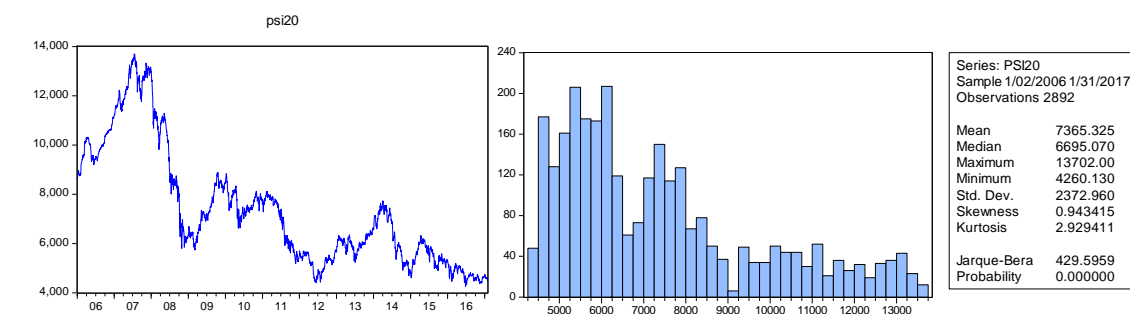


Figura 3.1. Análise Gráfica e Histograma da série PSI-20

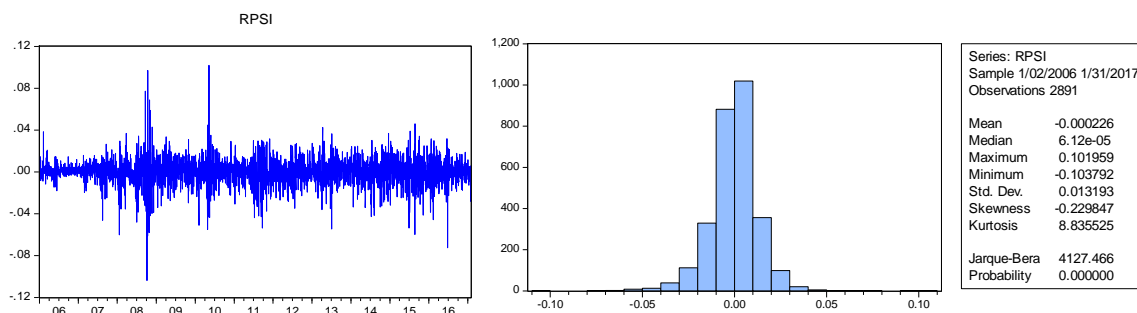


Figura 3.2. Análise Gráfica e Histograma da série dos retornos do PSI-20

Relembrar que o Jarque-Bera é um teste que pode ser usado para testar se as séries de retorno seguem uma distribuição gaussiana, com base nas observações de *skewness* e de *kurtosis* da distribuição empírica. Para seguir uma distribuição normal a *skewness* terá que ter um valor aproximado de 0 e a *kurtosis* um valor aproximado de 3.

Neste contexto, a série das rendibilidades de PSI-20 da figura 3.2 pode ser descrita como tendo uma média nula e um coeficiente de assimetria (*skewness*) próximo de zero, contudo o coeficiente de *kurtosis* apresenta um valor bastante elevado (8.8355), o que nos leva a concluir que a série não segue uma distribuição normal e que será precisa uma distribuição de caudas gordas para modelizar a leptocurtose presente.

A não-estacionaridade estuda-se em termos estatísticos pelos testes de raiz unitária, e como tal, para comprovar a não-estacionaridade da série do PSI-20 e a estacionaridade da série dos retornos, que foi observada graficamente, foram aplicados dois testes de raiz unitária, que já tinham sido apresentados, o teste de *Augmented Dickey-Fuller (ADF)* e o teste de *Phillips – Perron (PP)*.

Considere-se que a hipótese nula é de que exista uma raiz unitária, e portanto não seja estacionária, assim:

$H_0$  : existe uma raiz unitária (não é estacionária)

$H_1$ : não existe uma raiz unitária (é estacionária)

Os resultados obtidos (no *software Eviews*), em particular, os *p-values*, foram os seguintes,

Tabela 3.1. Testes de raiz unitária

<i>P-values</i>	ADF	PP
Série PSI-20	0.2722	0.2848
Série dos retornos	0.0001	0.0001

Aplicando os testes mencionados em cima (ADF e PP) verifica-se na tabela 3.1, que para a série PSI-20, para ambos os testes, as probabilidades são superiores a um nível de significância de 5% (ou 1%, ou 10%), não se rejeitando assim  $H_0$ . Portanto, a série apresenta a existência de uma raiz unitária, ou seja, é uma série não estacionária. Exatamente o contrário se verifica para a série dos retornos, para ambos os testes, registou-se um *P-value* de 0.0001. Como  $0.0001 < 0.05$  rejeita-se  $H_0$ , o que significa que para a série dos retornos logaritmizados, não existe uma raiz unitária, e por isso, é estacionária.

### 3.2 Estimação do VaR pelos modelos paramétricos e não-paramétricos

É de relembrar que existem, essencialmente, duas maneiras de calcular o VaR: não paramétrico e paramétrico.

Um dos métodos não-paramétrico é conhecido como Simulação Histórica (HS), a estimação do VaR é muito fácil de implementar se os dados ou os retornos estiverem disponíveis. Não há necessidade de quaisquer pressupostos quanto à forma da distribuição dos retornos, apenas que sejam independentes e identicamente distribuídos (i.i.d).

Por outro lado, o VaR paramétrico assume que os retornos seguem uma certa distribuição, o que simplifica, significativamente, o cálculo se esta distribuição for da família paramétrica.

O VaR pode ser comparado para vários modelos, testando-o nos retornos realizados e fornecendo evidências sobre a precisão das previsões.

Nesta parte irá ser feita uma análise para encontrar um modelo que melhor se ajuste aos dados das séries dos retornos com o objetivo de encontrar aquele que possa dar a melhor previsão do risco de mercado.

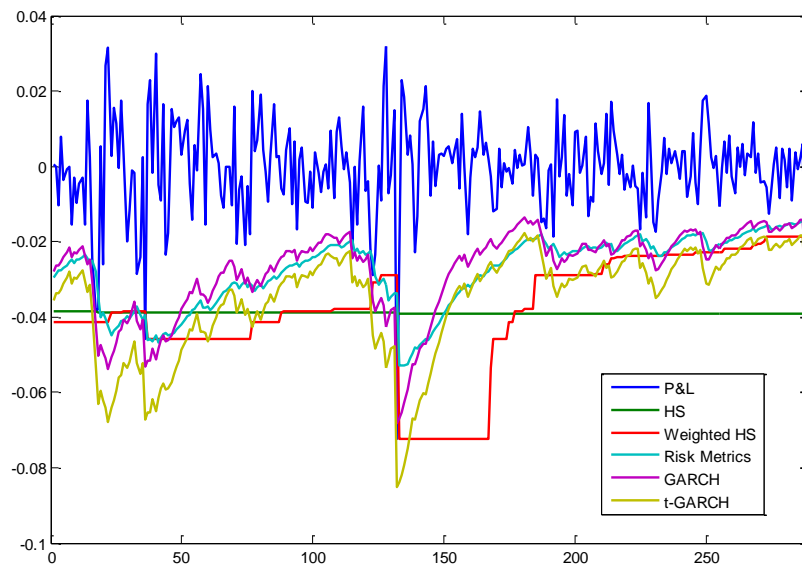


Figura 3.3. Output da estimação /previsão do VaR por vários métodos

Pela análise da figura 3.3 observa-se que um bom desempenho na avaliação do risco é gerado pelos modelos da família GARCH, tendo o pior desempenho os modelos não-paramétricos de tipo simulação histórica. Como tal, irá ser analisado com mais pormenor o *forecast* obtido pela utilização de dois modelos GARCH(1,1), o primeiro com distribuição normal e o segundo com distribuição *t-Student*.

A partir do software MATLAB implementaram-se modelos paramétricos da família GARCH para se obter previsões para as volatilidades e, assim, determinar valores para o VaR. Foram então estimados

os modelos GARCH (1,1) para as distribuições condicionadas Gaussiana e *t-Student*. Para a previsão do VaR usaram-se dois níveis de confiança: 99% (exigido pelo Comitê de Basileia) e 95% (utilizado pela metodologia *Riskmetrics*, que é uma caso particular de modelo GARCH). Relativamente ao horizonte temporal considerou-se uma janela temporal de tamanho 261 e considerou-se o VaR para 1 dia e para 5 dias (1 semana) para os níveis de significância de 99% e 95%.

A figura 3.4 apresenta os resultados do VaR (5 dias) nos modelos GARCH com distribuição normal e *t-Student*, para o nível de significância 99%. Observa-se que o risco estimado segue de forma bastante próxima o *trend* da série, mas tem um pequeno desfasamento em relação aos valores extremos.

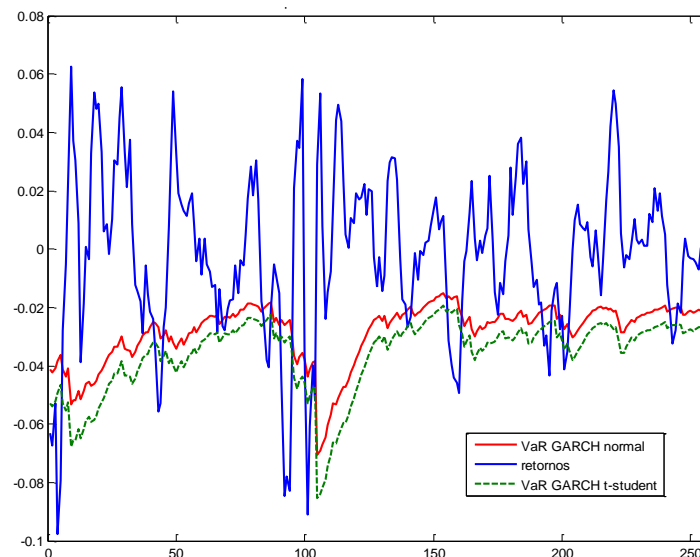


Figura 3.4. Análise gráfica para os modelos GARCH

As estimativas obtidas pelos modelos de risco devem ser avaliadas através de *backtesting* (referido anteriormente na secção 2.3), comparando as estimativas de risco com perdas efetivas por meio de testes estatísticos.

No final do *backtest*, o número de violações observadas é comparado ao número de violações previstas. Ou seja, após se proceder à análise da série de violações do VaR, compara-se, a taxa de rentabilidade da carteira para cada dia com a taxa de rentabilidade correspondente ao VaR, obtida na estimação para se obter o número de vezes em que a perda ocorrida ultrapassa o valor estimado para o VaR. Se este número for demasiado elevado então o modelo subestimou o risco, caso contrário, se este número for muito reduzido, ou até mesmo nulo, a medida de risco estimada é, provavelmente, demasiado alta. Se existirem grandes diferenças entre as violações, altas e baixas, então as medidas de ganhos e perdas poderão estar enviesadas.

Para a avaliação dos resultados dos modelos GARCH (1,1) foram considerados 2 testes, já referidos na secção 2.3.2, o teste de Kupiec (1995) e o teste de Christoffersen (1998). No entanto, vale a pena relembrar que Kupiec (1995) propôs um teste que permite verificar a proporção de vezes em que se verificam violações. Christoffersen (1998), por sua vez, apresentou uma extensão ao teste de Kupiec que avalia não só o número de violações mas também a independência entre elas.

Tabela 3.2. Resultados dos modelos na previsão do VaR para uma previsão de 5 dias

Teste Kupiec		Teste Christoffersen	
99%	95%	99%	95%

GARCH Normal	168.22	125.99	227.12	205.46
GARCH <i>t-Student</i>	106.08	90.04	157.40	169.99

Interpretando os testes de *backtesting*, de maneira a avaliar a adequabilidade e a *performance* dos modelos aplicados da estimação do VaR, para os níveis de confiança indicados e para uma previsão de 5 dias, pode-se concluir que segundo o teste de Kupiec tanto para as condições do modelo *Riskmetrics* como para o modelo Basileia rejeita-se a hipótese nula deste teste, pois, os valores obtidos da estatística de teste foram superiores ao valor crítico do qui-quadrado: 3.84.

Pelo teste de Christoffersen, verifica-se o mesmo, ambos os modelos não são considerados adequados, uma vez que todos apresentam valores da estatística de teste de Christoffersen superiores ao valor crítico do qui-quadrado: 5.99.

Assim, conclui-se que para uma previsão de 5 dias nenhum dos modelos GARCH para a previsão do VaR são validados pela *backtesting*.

Analisando os resultados obtidos para 1 dia de previsão (tabela 3.3) e, de acordo com as estatísticas de erro e níveis de confiança considerados, pode-se concluir que, ao contrário do que se verificou para uma previsão de 5 dias (tabela 3.2), todos dos modelos GARCH (1,1) do VaR são validados pela *backtesting*, obtendo-se uma boa previsão de risco.

Tabela 3.3. Resultados dos modelos na previsão do VaR para uma previsão de 1 dia

	Teste Kupiec		Teste Christoffersen	
	99%	95%	99%	95%
GARCH Normal	3.25	1.154	3.254	3.676
GARCH <i>t-Student</i>	0.0562	1.47	0.0562	2.528

Na tabela 3.4 observa-se que a percentagem de falha é menor para o modelo com distribuição *t-Student* e também para um nível de significância mais elevado (99%).

Tabela 3.4. Percentagem de falha

	Teste Christoffersen	
	99%	95%
GARCH Normal	0.023	0.065
GARCH <i>t-Student</i>	0.012	0.034

Para concluir, vejamos os valores preditos do VaR para 1 dia, com o nível de confiança de 99%, para os dois modelos validados acima. No caso GARCH com distribuição normal temos o valor 0.031 e no caso do modelo GARCH com distribuição *t-student* temos 0.039. Isto quer dizer que o valor da perda possível para 1 dia, não excede 3.1% e, respectivamente 3.9% do valor do stock com uma probabilidade de 99%.

### 3.3 Estimação do VaR pelos modelos EVT

As distribuições dos retornos de séries financeiras costumam apresentar caudas mais pesadas do que as de uma distribuição normal. Contudo, as metodologias tradicionais de cálculo do VaR, em geral, supõem normalidade e acomodam mal as ocorrências extremas da distribuição dos retornos, já referida na secção 2.4.

Entre os avanços recentes para os métodos do cálculo do VaR está a Teoria de Valores Extremos (EVT). A aplicação desta teoria permite estimar probabilidades e quantis com razoável precisão nos limites dos valores encontrados dentro da amostra, e nas caudas da distribuição.

Nesta secção, também com o auxílio do software MATLAB, irá ser feita uma análise do VaR para condições de valores extremos, para tal serão considerados os dois métodos mencionados ao longo da tese, o método dos blocos de máximos e o método *peak over threshold* (POT).

Recorde-se que o *block maxima method* é o método tradicional usado para analisar dados com sazonalidade. No entanto, o POT usa os dados de forma mais eficiente e, por essa razão, parece tornar-se o método mais usado para aplicações recentes. Este último considera a distribuição de excedências num certo limiar.

Numa primeira análise, os dados foram ajustados a uma distribuição generalizada de Pareto (GPD). Note-se que, quando os dados são traçados graficamente por uma distribuição exponencial ( $\xi = 0$ ), os desvios côncavos em torno de uma linha reta no gráfico são interpretados como indicação de distribuição de cauda pesada, enquanto que as distâncias convexas são interpretadas como uma indicação de uma cauda fina.

Através da observação do gráfico Q-Q (figura 3.5) verifica-se uma indicação gráfica de concavidade, o que é indicio de caudas pesadas na distribuição empírica dos retornos.

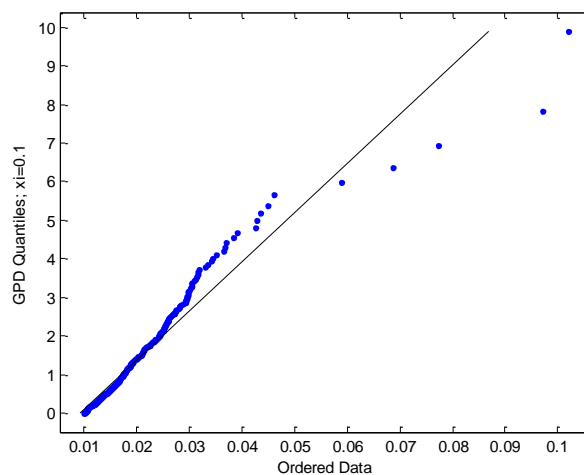


Figura 3.5. Análise gráfica da GPD

Como já tinha sido referido (secção 2.4.1), a aplicação do método dos blocos de máximos passa pelas seguintes etapas: dividir a amostra em  $n$  blocos de igual comprimento, seleccionar o valor máximo em cada bloco, ajustar a distribuição de GEV ao conjunto de máximos e, finalmente, estimar o ponto de cálculo e intervalo para  $R_n^k$ .

Para a análise dividiu-se a amostra em 10 blocos (anos) e foram encontrados 290 pontos de máx/min (figura 3.6).



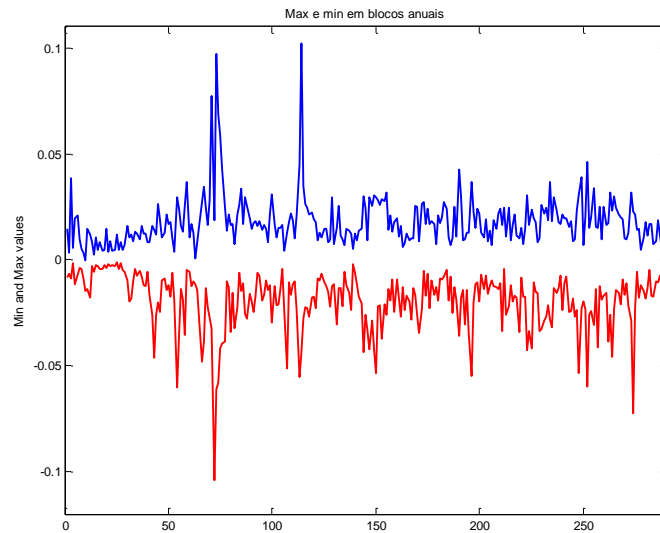


Figura 3.6. Análise gráfica da distribuição dos valores máximos

Após a aplicação do método do *block maxima*, para se verificar se o modelo é adequado foi feita uma análise aos resíduos.

Como se pode observar na figura 3.7 a distribuição dos resíduos não aproxima muito bem a exponencial, portanto o tamanho dos blocos anuais não é a melhor escolha para a distribuição GEV (*generalized extreme value*). Para o comprimento mensal dos blocos, os resultados são ainda mais fracos. Estes resultados devem ser consequências do número relativamente pequeno de dados da amostra, que representa uma das fraquezas deste modelo.

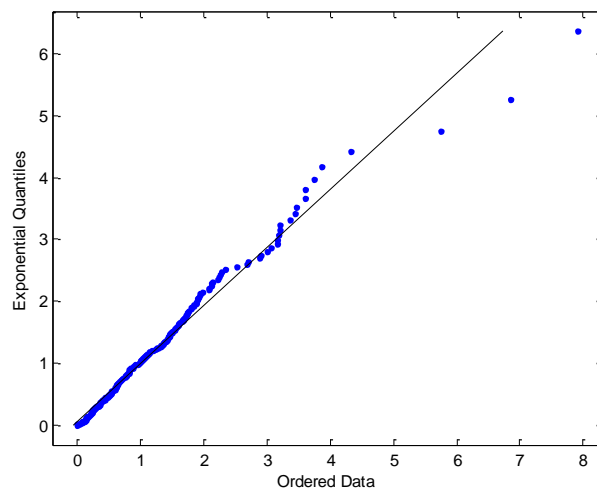


Figura 3.7. Análise gráfica dos resíduos

Como o método do *block maxima* não apresentou os melhores resultados, espera-se que o segundo método da teoria dos valores extremos, o POT, seja mais eficiente.

A implementação do método de POT envolve as seguintes etapas: seleccionar o *threshold*  $u$ , ajustar a função GPD para as excedências sobre  $u$  e, em seguida, calcular as estimativas de ponto e intervalo para *Value-at-Risk* (VaR). Assim, sabe-se que é muito importante a escolha do valor do *threshold*, pois, todos os valores acima desse *threshold* são ajustados na função GPD a partir da qual se obtém o VaR.

Para a determinação do *threshold* usou-se o *mean excess plot*, que se representa usando a função GPD (figura 3.8).

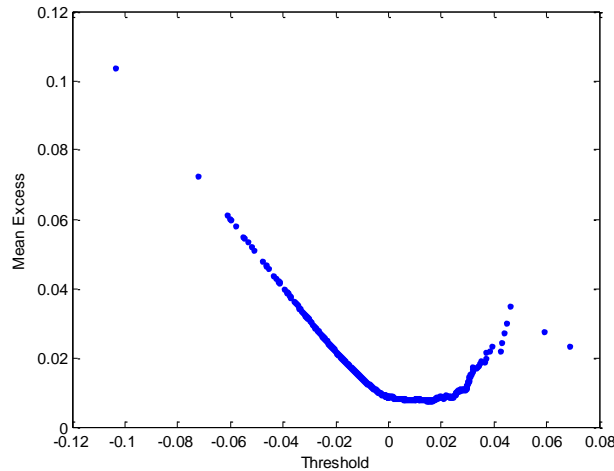


Figura 3.8. Mean excesso plot

A tendência linear ascendente indica um valor positivo para o parâmetro  $\xi$  (aproximadamente 0.16) e um valor aproximado de 0.0222 para o *threshold*.

Outra maneira de determinar o *threshold* é a partir da função Hill, este é um bom estimador por ser, por um lado, um estimador natural e, por outro, pela sua importância histórica.

A figura 3.9 representa a função que traça a estimativa Hill do índice da cauda contra as estatísticas da ordem superior  $k$  (número de excedências) ou contra diferentes limiares. O *threshold* é selecionado a partir do gráfico da figura 3.9 onde o parâmetro de forma ( $\xi$ ) é bastante estável. Pelo gráfico retira-se que o *threshold* da função Hill é 0.0222, corresponde à média dos pontos do intervalo de estabilidade. O que vai de acordo ao *threshold* determinado pelo *mean excess plot*.

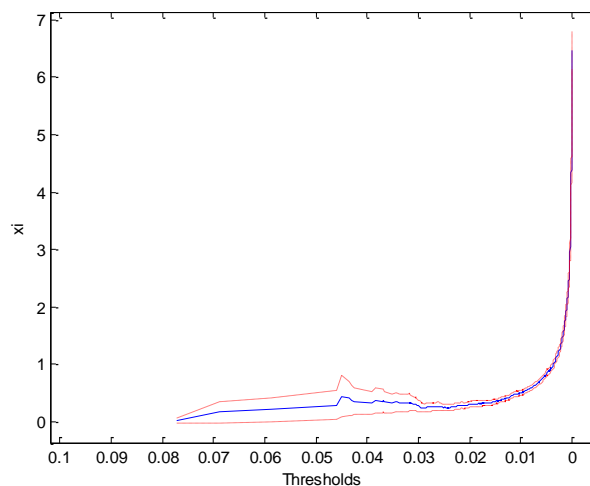


Figura 3.9. Gráfico da função Hill

A seguinte etapa deste método é ajustar a função GPD para as excedências sobre  $u$ . Assim, figura 3.10 representa o modelo GPD ajustado com as excedências acima do valor de *threshold* para a cauda direita do PSI-20. O gráfico sugere que o GPD estimado ajusta bastante bem as excedências.

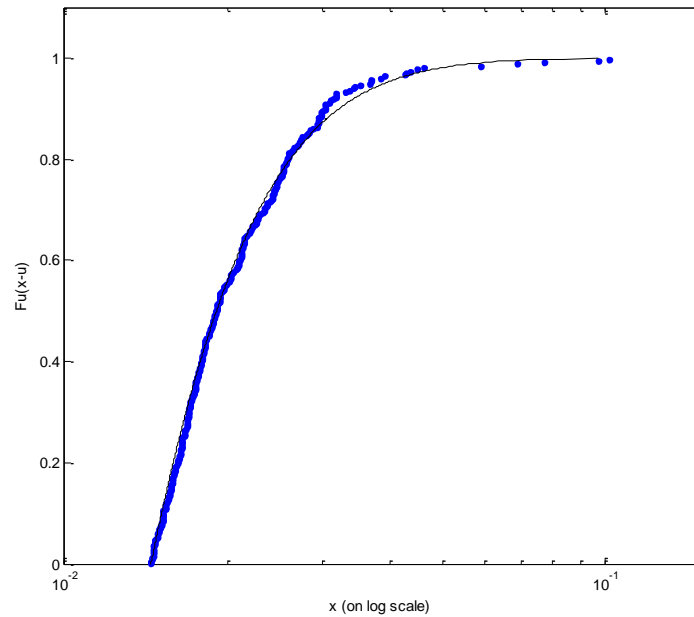


Figura 3.10. Modelo GPD ajustado com as excedências acima do valor de *threshold*

Finalmente, aplicou-se o modelo GPD, para o *threshold* 0.0222, e obteve-se os seguintes valores estimados para os parâmetros  $\beta = 0.0063$  e  $\xi = 0.1681$  (estimados pelo método da máxima verossimilhança). Como o  $\xi$  é positivo, tem-se mais um indício que a distribuição considerada é adequada para modelar distribuições de retornos financeiros com caudas pesadas.

Tal como para o método do *block maxima*, para o método POT também tem que se verificar se o modelo é adequado, e para tal foi feita uma análise aos resíduos. Como se pode observar na figura 3.11 a distribuição dos resíduos aproxima bastante bem a exponencial, ligeiramente melhor do que o resultado para o outro método, portanto o modelo tem um bom desempenho, como pode ser observado no Q-Q *plot* dos resíduos.

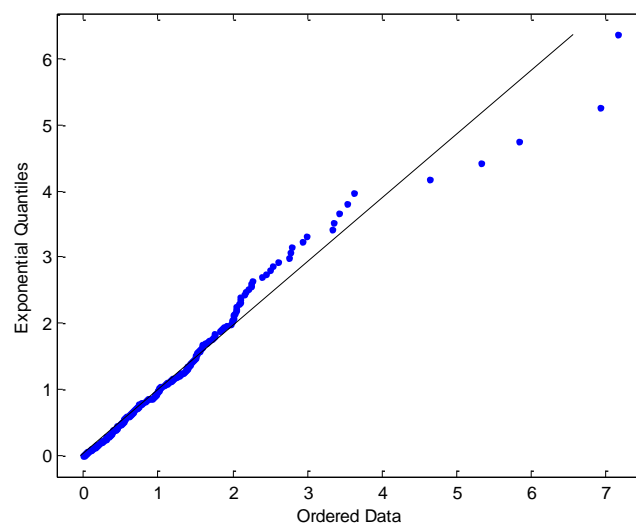


Figura 3.11. Q-Q *plot* dos resíduos

Uma vez que a distribuição dos excessos sobre um *threshold* está estimada, uma aproximação da distribuição original e uma estimativa do  $p$ -quantil podem ser usadas para estimar o VaR extremo.

Vamos considerar a previsão do VaR para um dia, pois, para este intervalo temporal obtivemos os melhores resultados nos outros modelos.

A figura 3.12 mostra que a estimação do quantil é de 1%, ou seja, o VaR é de 1%. Conclui-se assim que, para um nível de confiança de 99%, podemos prever as perdas de amanhã (cauda direita) do PSI-20, sem exceder 5.84% do valor do *stock*.

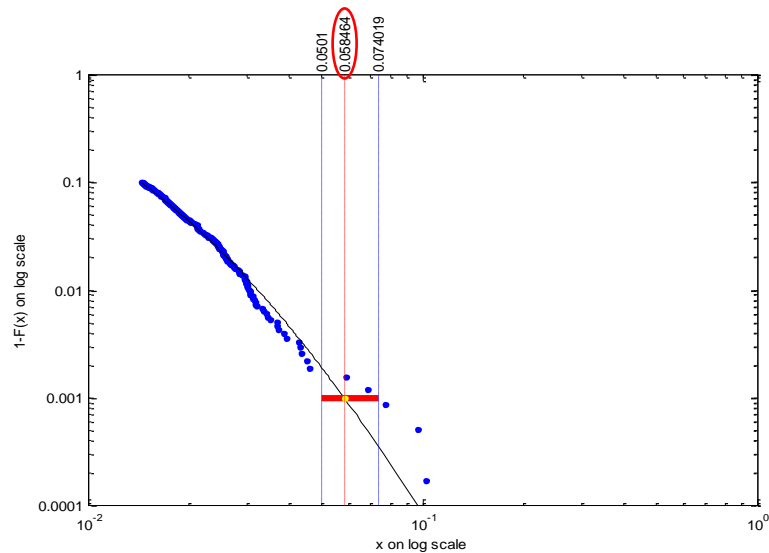


Figura 3.12. Estimação do quantil

Para concluir este capítulo, temos que os modelos GARCH determinam uma melhor previsão do VaR do que os modelos baseados na teoria dos valores extremos. O resultado pode ser justificado pelo número reduzido de valores extremos na amostra considerada. Os resultados podem ser melhorados pela consideração de vários valores de *threshold*, níveis de significância diferentes ou pela ampliação da própria amostra. Tudo isto pode fazer parte de um próximo trabalho, pois aqui, limitamos o nosso objectivo a um exercício de aplicação dessas metodologias.

## CONCLUSÃO

Desde cedo, as Finanças, têm procurado medir o nível de risco envolvido nos diversos tipos de investimentos. Inicialmente, usou-se a volatilidade (variância), mas a necessidade de monetarizar o risco levou à utilização do VaR (*Value-at-Risk*). No entanto estas medidas não são consideradas medidas de risco coerentes. Esta situação está a levar reguladores e instituições financeiras a direccionar os seus esforços para uma medida mais rigorosa o *Expected Shortfall* (ES) que além de ser uma medida coerente é também uma medida de risco consistente. Contudo nenhuma destas medidas conseguia prever os acontecimentos de valores extremos.

Ao longo deste trabalho tentou-se demonstrar a utilidade da Teoria de Valores Extremos, no cálculo da medição do risco (VaR), no que respeita à análise da cauda das distribuições, onde a simulação histórica se revela insuficiente na análise de eventos extremos.

Utilizando a cotação do PSI-20 foram feitas aplicações práticas dos modelos anteriores, com o objetivo de verificar se são ou não adequados.

Depois de ter sido feita a análise dos dados e dos resultados, concluiu-se que os modelos apresentam desempenhos diferentes na previsão do VaR. Para avaliar o desempenho dos modelos foram efetuados os testes de *backtesting* de Kupiec e de Christoffersen. Para tal, foi analisado com mais pormenor o *forecast* obtido pela utilização de dois modelos GARCH(1,1), o primeiro com distribuição normal e o segundo com distribuição *t-Student*.

Relativamente ao horizonte temporal considerou-se uma janela temporal de tamanho 261 e considerou-se o VaR para 1 dia e para 5 dias (1 semana) para os níveis de significância de 99% e 95%.

A aplicação prática destes modelos, na série considerada, permitiu concluir que para uma previsão de 5 dias os dois modelos GARCH considerados do VaR não são validados pela *backtesting*. Enquanto que, para 1 dia de previsão todos os modelos GARCH (1,1) do VaR são validados pela *backtesting*, obtendo-se uma boa previsão de risco.

Para a estimação do VaR considerando os modelos da Teoria dos Valores Extremos, foi testados dois métodos, o método do *Block Maxima* e o método do *peak over threshold* (POT). Para o primeiro método os resultados não foram muito satisfatórios, tendo-se verificado que o tamanho dos blocos anuais não é a melhor escolha para a distribuição GEV (*generalized extreme value*). Enquanto que o método POT comprovou ser um método mais eficiente, tendo o modelo apresentado um bom desempenho nos resultados.

Os resultados obtidos indicam que o uso da EVT para o cálculo do VaR é apropriado para o método POT, enquanto que o uso da modelação do tipo GARCH para os níveis de significância de 95% e 99% é bem indicado para uma previsão de 1 dia.

Foi feita uma análise empírica para se avaliar o desempenho de modelos VaR alternativos. O resultado foi que a EVT parece ser uma alternativa razoável quando não se conhece a distribuição dos retornos, mas, neste caso particular, o VaR estimado pelos modelos GARCH tem um melhor desempenho.

As abordagens apresentadas encontram-se ainda em fase de desenvolvimento. Por isso, é neste âmbito que para percursos futuros poderá ser interessante tentar aplicar-se novos testes de *backtesting*. Fica também como sugestão a realização de um trabalho semelhante para carteiras de ativos ou de investimentos, e verificar se o fato de se agrupar um maior conjunto de ativos ou investimentos afeta de alguma forma os resultados obtidos com o uso dessas técnicas mais sofisticadas.

## BIBLIOGRAFIA

- Adams, M., & Thornton, B. (s.d.). Black swans and VaR. *Journal of Finance and Accountancy*.
- Aeppli, M. D. (2011). On the Design of Stress Tests. UNIVERSITY OF ST. GALLEN: Master of Arts in Banking and Finance.
- Alexandar, C. (. (2008). *Market Risk Analysis: Practical Financial Econometrics (Vol. II)*. Wiley Publishing.
- Allen, D. E., Singh, A. K., & Powell, R. J. (2011). Extreme Market Risk - An Extreme Value Theory Approach. *ECU Publications 2011*. Edith Cowan University.
- Andersen, T. T., & Frederiksen, T. C. (2010). *Modelling Value-at-Risk under Normal and Extreme Market Conditions*. Copenhagen Business School: Masters Thesis.
- Christoffersen, P. F., Diebold, F., & Schuermann, T. (1998). Horizon problems and extreme events in financial risk management. *Economic Policy Review(Oct)*, 109-118.
- Coles, S. G. (2001). *An introduction to Statistical Modelling of Extreme Values*. Verlag: Springer.
- Dacorogna, M. M., Müller, U. A., Pictet, O. V., & Vries, C. G. (2001). Extremal Forex Returns in Extremely Large Data Sets. *Extremes*, 4(2), pp. 105-127.
- Danielsson, J. (2002). The emperor has no clothes: Limits to risk modelling. *Journal of Banking & Finance*, 26(7), 1273.
- Danielsson, J. (2011). *Financial risk forecasting: The theory and practice of forecasting market risk, with implementation in R and MATLAB*. Chichester: John Wiley.
- DANIELSSON, J. H. (2000). *Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation*. Econometric Institute Report EI 2000-19/A.
- Danielsson, J., & de Vries, C. (2000). Value-at-Risk and Extreme Returns. *Annales d'Economie et de Statistique*, 60, pp. 239-270.
- Dupuis, D. J. (1998). Robust estimation of extremes. *The Canadian Journal of Statistics*.
- Embrechts, P. K. (1997). *Modelling extremal events for insurance and finance*. Springer, Berlin.
- Gençay, R., Selçuk, F., & Ulugulyagc, A. (2001). EVIM: A Software Package for Extreme Value Analysis in Matlab. *Studies in Nonlinear Dynamics and Econometrics*, (pp. 213-239).
- Gilli, M. &. (2006). *An Application of Extreme Value Theory for Measuring Financial Risk*. Computational Economics.
- Gilli, M., & Kellezi, E. (2006). An Application of Extreme Value Theory for Measuring Financial Risk. *Computational Economics*, (pp. 1-23).
- GILLOT, P., & PION, D. (1993). *Le Nouveau Cambisme*. (E. ESKA, Ed.) Paris.
- Holton, G. (2003). *Value-at-Risk: Theory and Practice: Academic Press*.

- Holton, G. A. (2004). Defining Risk . *Financial Analysts Journal*.
- Izmaylov, B. (2014). *Value-at-Risk: Strengths, Caveats and Considerations for Risk Managers and Regulators*. Aarhus University: Master Thesis.
- Jorion, P. (2007). *Value at Risk: The new benchmark for managing financial risk (3. edition ed.)*. London: McGraw-Hill.
- Koedijk, K. G., Schafgans, M., & Vries, C. (1990). The Tail Index of Exchange Rate Returns. *Journal of International Economics*, 29, 93-108.
- Loretan, M., & Phillips, P. (1994). Testing the covariance stationarity of heavy-tailed time series. *Journal of Empirical Finance*, 1 (2), 211-248.
- Martins, V. M. (2015). *Expected Shortfall: Algumas abordagens de implementação em séries financeiras*. Dissertação do Mestrado de Matemática Financeira.
- McNeil, A. J. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3-4):271-300.
- McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.
- Mehta, A., Neukirchen, M., Pfetsch, S., & Poppensieker, T. (2012). *Managing market risk: Today and tomorrow*. McKinsey working papers on risk, no. 32.
- Neftci, S. N. (2000). Value at risk calculations, extreme events, and tail estimation. *Journal of Derivatives*, 23-37.
- Nystrom, K. &. (2002). *Univariate Extreme Value Theory, GARCH and Measures of Risk*. Preprint, Swedbank.
- Pritsker, M. (1997). Evaluating value at risk methodologies: Accuracy versus computational time. *Journal of Financial Services Research*, 12(2), 201.
- Taleb, N. (2007). *Fooled by randomness*. Obtido de <http://www.fooledbyrandomness.com/imbeciles.htm>
- Vicente, S. L. (2012). *EXTREME VALUE THEORY: AN APPLICATION TO SPORTS*. Dissertação do Mestrado em Estatística.



## ANEXOS

### Anexo 1: Teste de raiz unitária à série PSI-20 para o teste ADF

Null Hypothesis: PSI20 has a unit root				
Exogenous: None				
Lag Length: 1 (Automatic - based on SIC, maxlag=27)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-1.020034	0.2772
Test critical values:	1% level		-2.565762	
	5% level		-1.940933	
	10% level		-1.616626	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(PSI20)				
Method: Least Squares				
Date: 08/17/17 Time: 13:56				
Sample (adjusted): 1/04/2006 1/31/2017				
Included observations: 2890 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
PSI20(-1)	-0.000227	0.000223	-1.020034	0.3078
D(PSI20(-1))	0.064775	0.018563	3.489530	0.0005
R-squared	0.004327	Mean dependent var		-1.460941
Adjusted R-squared	0.003982	S.D. dependent var		92.85825
S.E. of regression	92.67319	Akaike info criterion		11.89673
Sum squared resid	24803069	Schwarz criterion		11.90086
Log likelihood	-17188.77	Hannan-Quinn criter.		11.89822
Durbin-Watson stat	2.001098			

**Anexo 2: Teste de raiz unitária à série PSI-20 para o teste PP**

Null Hypothesis: PSI20 has a unit root				
Exogenous: None				
Bandwidth: 8 (Newey-West automatic) using Bartlett kernel				
		Adj. t-Stat	Prob.*	
<hr/>				
Phillips-Perron test statistic		-1.000912	0.2848	
Test critical values:	1% level	-2.565762		
	5% level	-1.940933		
	10% level	-1.616626		
<hr/>				
*MacKinnon (1996) one-sided p-values.				
<hr/>				
Residual variance (no correction)			8618.843	
HAC corrected variance (Bartlett kernel)			9676.975	
<hr/>				
Phillips-Perron Test Equation				
Dependent Variable: D(PSI20)				
Method: Least Squares				
Date: 08/17/17 Time: 13:57				
Sample (adjusted): 1/03/2006 1/31/2017				
Included observations: 2891 after adjustments				
<hr/>				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
<hr/>				
PSI20(-1)	-0.000228	0.000223	-1.020980	0.3073
<hr/>				
R-squared	0.000124	Mean dependent var	-1.427537	
Adjusted R-squared	0.000124	S.D. dependent var	92.85955	
S.E. of regression	92.85378	Akaike info criterion	11.90027	
Sum squared resid	24917074	Schwarz criterion	11.90234	
Log likelihood	-17200.85	Hannan-Quinn criter.	11.90102	
Durbin-Watson stat	1.870070			
<hr/>				

**Anexo 3: Teste de raiz unitária à série dos retornos para o teste ADF**

Null Hypothesis: RPSI has a unit root

Exogenous: None

Lag Length: 0 (Automatic - based on SIC, maxlag=27)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-49.48728	0.0001
Test critical values: 1% level	-2.565762	
5% level	-1.940933	
10% level	-1.616626	

\*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(RPSI)

Method: Least Squares

Date: 08/17/17 Time: 13:59

Sample (adjusted): 1/04/2006 1/31/2017

Included observations: 2890 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RPSI(-1)	-0.917452	0.018539	-49.48728	0.0000
R-squared	0.458785	Mean dependent var		-4.29E-06
Adjusted R-squared	0.458785	S.D. dependent var		0.017875
S.E. of regression	0.013150	Akaike info criterion		-5.824411
Sum squared resid	0.499590	Schwarz criterion		-5.822345
Log likelihood	8417.273	Hannan-Quinn criter.		-5.823666
Durbin-Watson stat	1.998808			

**Anexo 4: Teste de raiz unitária à série dos retornos para o teste PP**

Null Hypothesis: RPSI has a unit root

Exogenous: None

Bandwidth: 7 (Newey-West automatic) using Bartlett kernel

	Adj. t-Stat	Prob.*
Phillips-Perron test statistic	-49.35489	0.0001
Test critical values: 1% level	-2.565762	
5% level	-1.940933	
10% level	-1.616626	

\*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	0.000173
HAC corrected variance (Bartlett kernel)	0.000158

Phillips-Perron Test Equation

Dependent Variable: D(RPSI)

Method: Least Squares

Date: 08/17/17 Time: 13:58

Sample (adjusted): 1/04/2006 1/31/2017

Included observations: 2890 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RPSI(-1)	-0.917452	0.018539	-49.48728	0.0000
R-squared	0.458785	Mean dependent var	-4.29E-06	
Adjusted R-squared	0.458785	S.D. dependent var	0.017875	
S.E. of regression	0.013150	Akaike info criterion	-5.824411	
Sum squared resid	0.499590	Schwarz criterion	-5.822345	
Log likelihood	8417.273	Hannan-Quinn criter.	-5.823666	
Durbin-Watson stat	1.998808			